Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance

Harrison H. Barrett

Department of Radiology and Optical Sciences Center, University of Arizona, Tucson, Arizona 85724

J. L. Denny

Department of Radiology and Department of Mathematics, University of Arizona, Tucson, Arizona 85724

Robert F. Wagner and Kyle J. Myers

Center for Devices and Radiological Health, U.S. Food and Drug Administration, Rockville, Maryland 20852

Received July 5, 1994; accepted December 8, 1994

Figures of merit for image quality are derived on the basis of the performance of mathematical observers on specific detection and estimation tasks. The tasks include detection of a known signal superimposed on a known background, detection of a known signal on a random background, estimation of Fourier coefficients of the object, and estimation of the integral of the object over a specified region of interest. The chosen observer for the detection tasks is the ideal linear discriminant, which we call the Hotelling observer. The figures of merit are based on the Fisher information matrix relevant to estimation of the Fourier coefficients and the closely related Fourier crosstalk matrix introduced earlier by Barrett and Gifford [Phys. Med. Biol. **39,** 451 (1994)]. A finite submatrix of the infinite Fisher information matrix is used to set Cramer–Rao lower bounds on the variances of the estimates of the first N Fourier coefficients. The figures of merit for detection tasks are shown to be closely related to the concepts of noise-equivalent quanta (NEQ) and generalized NEQ, originally derived for linear, shift-invariant imaging systems and stationary noise. Application of these results to the design of imaging systems is discussed.

1. INTRODUCTION

Many modern imaging systems are digital, in the sense that the data consist of a finite set of discrete measurements. Often this digital data set is not itself a useful image, and an image-reconstruction step is required for production of an image; tomography is a prime example of such indirect, digital imaging. Since the object being imaged is a continuous function, the digital imaging system is a continuous-to-discrete mapping,[1,2] and the data set is necessarily an incomplete representation of the object. A crucial question is how to design the imaging system in such a way that a useful image, with a certain spatial resolution and relative freedom from artifacts, can be reconstructed.

Phrased in this way the question is rather vague, and more objective definitions of image quality are needed. There is considerable interest, especially in the medical-imaging literature, in defining image quality in terms of how well some observer can perform some specified task of practical importance.[3–17] Generically, the tasks can be either to classify the image into one of two or more categories or to estimate one or more quantitative parameters of the object from the image data. We call these two kinds of task classification and estimation, respectively. The observer can be either a human or some mathematical model observer such as the Bayesian ideal observer. For a survey of figures of merit for task performance by various model observers, see the previous paper in this series,[18] referred to hereafter as paper I.

The simplest classification task is detection of a known signal on a known background: signal known exactly, background known exactly (SKE/BKE). For both human and ideal observers, performance on this detection task is limited primarily by noise in the data or in the reconstructed image, and deterministic properties such as spatial resolution and artifacts are relatively unimportant.[19,20] More complicated classification tasks as well as many estimation tasks place further demands on the deterministic properties.[9,10,20–24]

In paper I, figures of merit for both classification and estimation tasks were derived. It was shown that an imaging system optimized for a certain classification task was not necessarily optimized for other classification tasks or for estimation tasks. A number of relations among figures of merit for different tasks and observers were derived, but no general strategy for design and optimization of imaging systems was given. In paper I the object to be imaged was represented as a discrete set of voxels, but the voxel size was allowed to be very small, and system null functions were explicitly considered.

In a more recent paper Barrett and Gifford (BG hereafter) suggested a strategy for design of cone–beam tomography systems.[25] In that paper a continuous object of compact support was represented exactly by a Fourier series multiplied by a zero-one function for enforcement of the support constraint. No specific task was considered, and the purpose of the imaging system was assumed to be to provide data from which as many as possible of the Fourier coefficients could be estimated. To quantify the ability of the system to provide this information, BG introduced a concept called the crosstalk matrix, defined in Section 2 below. This matrix is a measure of how well each Fourier coefficient is recoverable from the discrete data set. The diagonal elements of this matrix relate to a generalized transfer function, specifying the strength of a Fourier component as reflected in the data, while the off-diagonal elements give the degree of linear dependence or aliasing of two different components. The design strategy suggested in BG was to make the crosstalk matrix as nearly diagonal as possible.

It is not obvious how to relate the BG design strategy, or any other approach that focuses on the deterministic properties of the image, to performance of specific estimation or classification tasks. One possible unifying concept is that of Fisher information. It is well known that a bound on the performance of an estimator can be set if the Fisher information matrix is known[26–28] and that this bound is attainable in many cases of practical interest. Thus it is not unexpected that system performance for an estimation task can be related to the Fisher matrix, and it is also known that a Fisher information matrix enters into the figure of merit for performance of classifications tasks.[26] It is important to note, however, that there are many different Fisher information matrices, depending on what parameters one wants to estimate. It is therefore not very useful to say that one should optimize some measure derived from the Fisher information matrix without specifying which Fisher information matrix and which task and observer are meant.

It is the goal of this paper to tie together the concepts of Fourier crosstalk, Fisher information, and task performance, including classification tasks more complicated than SKE/BKE, as well as estimation tasks. As in BG the imaging system is treated in full generality as a linear, continuous-to-discrete mapping. We show that the crosstalk matrix, though treated purely deterministically in BG, also arises naturally when noise is considered. It is an approximation to the Fisher information matrix for estimation of the Fourier coefficients under certain assumptions about the noise, and in one important limit the Fisher and the crosstalk matrices differ by a constant. With these noise assumptions we are able to express the various figures of merit for image quality in terms of the crosstalk matrix and relate the BG design strategy to optimization of image quality as defined objectively by specific tasks.

We also show how the crosstalk matrix is related to the concepts of noise-equivalent quanta (NEQ) and generalized NEQ. For linear, shift-invariant imaging systems and stationary noise, the NEQ approach leads to compact Fourier-domain expressions for classification figures of merit, but neither of these assumptions is likely to be valid for tomographic or other indirect imaging systems. Nevertheless, we show that a similar analysis applies much more generally and that expressions analogous to NEQ can be derived without assuming shift invariance or noise stationarity. Perhaps surprisingly, the Fourier domain turns out to be extremely useful even for shift-variant imaging systems and nonstationary noise.

Though the formalism developed in paper I includes properties of both the imaging system and the reconstruction or processing algorithm, the focus in the present paper is almost exclusively on system design. Indeed, some of the figures of merit will be derived on the assumption that a mathematical observer has access to the raw data without any reconstruction algorithm. When a reconstruction step is invoked, the algorithm will be assumed to be optimal in the sense that it delivers efficient, unbiased estimates of the object Fourier coefficients. With this restriction, however, the results are general enough to apply to a wide variety of imaging systems, including shadow-casting systems as in radiology and systems in which the wave properties of the radiation come into play.

2. CROSSTALK CONCEPT

A. Object Representation

We consider an object described by a density function $f(\mathbf{r})$, where $\mathbf{r}$ is a three-dimensional (3D) vector, but the results are easily modified for two-dimensional (2D) and one-dimensional (1D) objects. We assume that $f(\mathbf{r})$ is zero outside some finite spatial region such as a cube or sphere and define a support function $S(\mathbf{r})$, which is unity when $\mathbf{r}$ is inside this region and zero otherwise. As a result of the finite support, the object can be represented exactly by the Fourier series

$$f(\mathbf{r}) = \sum_{k=1}^{\infty} F_k \mathbf{\Phi}_k(\mathbf{r}), \tag{1}$$

where

$$\Phi_k(\mathbf{r}) = \exp(2\pi i \boldsymbol{\rho}_k \cdot \mathbf{r}) S(\mathbf{r}). \tag{2}$$

Each basis function $\Phi_k(\mathbf{r})$ in this series describes a plane wave truncated by the support function. The wave vectors $\{\boldsymbol{\rho}_k\}$ are chosen to define points on an infinite 3D cubic lattice. Three integer indices, each running from $-\infty$ to ∞ are required for specification of this lattice in Fourier space, but for notational simplicity we lump them into a single index k that enumerates points on the lattice. We assume that the Fourier coefficients are ordered by increasing spatial frequency, so that $k > k'$ implies $|\boldsymbol{\rho}_k| \geq |\boldsymbol{\rho}_{k'}|$. In addition, since $f(\mathbf{r})$ is real, $F_k = F_{k'}^*$ if $\boldsymbol{\rho}_k = -\boldsymbol{\rho}_{k'}$. Sums with an infinite upper limit are assumed to run over this infinite 3D lattice in Fourier space, while finite sums, $k = 1 \ldots N$, include the N smallest values of $|\boldsymbol{\rho}_k|$ but always include all complex-conjugate pairs so that the sum is real. This representation of $f(\mathbf{r})$ is exact if the spacing of points on the 3D lattice satisfies a Nyquist condition related to the size of the support[25] and the full infinite sum is retained.

B. Representation of the Imaging System

The discrete data obtained from this continuous object are assumed to consist of M measurements g_m, $m = 1 \ldots M$, each of which is a random variable. We assume that the measurement system is linear, so that the mean of the mth measurement, denoted $\overline{g}_m$, is related to $f(\mathbf{r})$ by

$$\overline{g}_m = \int_S f(\mathbf{r}) h_m(\mathbf{r}) \mathrm{d}\mathbf{r}\,, \tag{3}$$

where $h_m(\mathbf{r})$ is called the detector sensitivity function because it describes the average sensitivity of the mth measurement to the object density at point $\mathbf{r}$. The integral runs over the region of support of the object as defined by $S(\mathbf{r})$. Note that $h_m(\mathbf{r})$ is linearly proportional to the integration time of the measurement. Thus, if $\overline{g}_m$ is the mean number of photons detected in time τ, it is proportional to τ, and this factor is assumed to be contained in $h_m(\mathbf{r})$.

Noise in the data arises from counting statistics or excess noise in the detector or electronics. To describe the noise in the mth measurement, we define $\epsilon_m \equiv g_m - \overline{g}_m$, where the overbar denotes an ensemble average over noise realizations for a fixed (nonrandom) object. All three quantities in this expression can be regarded as components of $M \times 1$ column vectors, $\boldsymbol{\epsilon}$, $\mathbf{g}$, and $\overline{\mathbf{g}}_m$, respectively, so we can write abstractly

$$\mathbf{g} = \overline{\mathbf{g}} + \boldsymbol{\epsilon} = \mathcal{H} f(\mathbf{r}) + \boldsymbol{\epsilon}\,, \tag{4}$$

where $\mathcal{H}$ is the continuous-to-discrete mapping defined by Eq. (3). Using the series of Eq. (1) to represent $f(\mathbf{r})$, we have

$$\mathbf{g} = \mathcal{H} \sum_{k=1}^{\infty} F_k \Phi_k(\mathbf{r}) + \boldsymbol{\epsilon} = \sum_{k=1}^{\infty} F_k \mathcal{H} \Phi_k(\mathbf{r}) + \boldsymbol{\epsilon}\,. \tag{5}$$

Throughout this paper we use m to denote a specific detector and k or k' to denote a Fourier component.

In component form, Eq. (5) is

$$g_m = \sum_{k=1}^{\infty} F_k [\mathcal{H} \Phi_k(\mathbf{r})]_m + \epsilon_m \equiv \sum_{k=1}^{\infty} \Psi_{mk} F_k + \epsilon_m\,. \tag{6}$$

The quantity Ψ_{mk}, given explicitly by

$$\Psi_{mk} = [\mathcal{H} \Phi_k(\mathbf{r})]_m = \int_S h_m(\mathbf{r}) \exp(2\pi i \boldsymbol{\rho}_k \cdot \mathbf{r}) \mathrm{d}\mathbf{r}\,, \tag{7}$$

is the Fourier *transform* of the product of the sensitivity function and the support function, with the transform evaluated at the lattice point $\boldsymbol{\rho}_k$. It can also be interpreted as the contribution of the kth Fourier basis function $\Phi_k(\mathbf{r})$ to the mth measurement. (See BG for an interpretation of Ψ_{mk} in terms of line integrals for the tomography problem.)

Thus the measured data and the Fourier coefficients are related by

$$\mathbf{g} = \boldsymbol{\Psi} \mathbf{F} + \boldsymbol{\epsilon}\,, \tag{8}$$

where the elements of the matrix $\boldsymbol{\Psi}$ are the values Ψ_{mk} ($m = 1 \ldots M$, $k = 1 \ldots \infty$). (Note that $\boldsymbol{\Psi}\mathbf{F}$ is real even though $\boldsymbol{\Psi}$ and $\mathbf{F}$ individually are complex.) Even though our object representation is still continuous, we have reduced the forward problem to matrix-vector form but with a matrix of dimension $M \times \infty$.

C. Crosstalk Matrix

There are two distinct problems in recording enough information to be able to recover a particular Fourier coefficient F_k from a discrete set of measurements. First, this Fourier component must make a significant contribution to $\mathbf{g}$. Second, this contribution must be distinguishable from the contribution made by other Fourier components. In familiar terms, the transfer function must be nonzero and aliasing should be avoided.

The crosstalk matrix is a way of quantifying these problems. This infinite, complex matrix is denoted $\mathbf{B}$, with elements defined by

$$\beta_{kk'} = \sum_{m=1}^{M} \Psi_{mk}^{*} \Psi_{mk'}\,. \tag{9}$$

An equivalent definition is

$$\mathbf{B} = \boldsymbol{\Psi}^{\dagger} \boldsymbol{\Psi}\,, \tag{10}$$

where $\boldsymbol{\Psi}^{\dagger}$ is the adjoint (transpose of the complex conjugate) of $\boldsymbol{\Psi}$. From either of these equations it is easy to show that $\mathbf{B}$ is Hermitian.

Equation (9) will also be recognized as the discrete scalar product of the mean data vector produced by the basis function $\Phi_k(\mathbf{r})$ with the mean data vector produced by $\Phi_{k'}(\mathbf{r})$; i.e.,

$$\begin{aligned} \beta_{kk'} &= \sum_{m=1}^{M} [\mathcal{H} \Phi_k(\mathbf{r})]_m^{*} [\mathcal{H} \Phi_{k'}(\mathbf{r})]_m \\ &= (\mathcal{H} \Phi_k(\mathbf{r}),\, \mathcal{H} \Phi_{k'}(\mathbf{r}))\,, \end{aligned} \tag{11}$$

where $(\mathbf{u}, \mathbf{v})$ denotes the complex scalar product of two M-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$.

The diagonal element of the crosstalk matrix is given by

$$\beta_{kk} = \sum_{m=1}^{M} |\Psi_{mk}|^2 = \sum_{m=1}^{M} \left| \int_S h_m(\mathbf{r}) \exp(2\pi i \boldsymbol{\rho}_k \cdot \mathbf{r}) \mathrm{d}\mathbf{r} \right|^2, \tag{12}$$

which is just the square of the L_2 norm of the $M \times 1$ vector $\mathcal{H} \Phi_k(\mathbf{r})$. The element β_{kk} thus measures the strength of the kth Fourier component in the data set; if β_{kk} is zero, the component makes no contribution to the data and cannot be recovered by any algorithm.

To see the significance of $\beta_{kk'}$ for $k \neq k'$, we use Eq. (11). If the vectors $\mathcal{H} \Phi_k(\mathbf{r})$ and $\mathcal{H} \Phi_{k'}(\mathbf{r})$ are orthogonal, then $\beta_{kk'} = 0$, and it is quite easy to devise algorithms to distinguish them. In general, the angle in the M-dimensional data space between $\mathcal{H} \Phi_k(\mathbf{r})$ and $\mathcal{H} \Phi_{k'}(\mathbf{r})$ can be defined by

$$\cos\,\theta_{kk'} = \sqrt{\frac{|\beta_{kk'}|^2}{\beta_{kk} \beta_{k'k'}}}\,. \tag{13}$$

A nonzero value for $\beta_{kk'}$ can thus be interpreted as meaning that the angle $\theta_{kk'}$ is less than $\pi/2$, which makes it harder to discriminate $\Phi_k(\mathbf{r})$ from $\Phi_{k'}(\mathbf{r})$. In the extreme case where $\theta_{kk'} = 0$, the basis functions $\Phi_k(\mathbf{r})$and $\Phi_{k'}(\mathbf{r})$ produce exactly the same pattern in the data and cannot be distinguished by any algorithm.

D. A Design Strategy

Since it is obviously impossible to estimate the entire infinite set of Fourier coefficients from a finite number of measurements, we choose some subset of the coefficients and attempt to estimate them. One reasonable choice is to attempt to estimate the N coefficients for which $|\boldsymbol{\rho}_k|$ is less than some maximum value; call it $\rho_{\max}$. We consider a finite $(N \times N)$ submatrix of $\mathbf{B}$, denoted $\mathbf{B}_N$ and defined in the same way as $\mathbf{B}$ but with k and k' restricted to the range $1 \ldots N$. For the N Fourier coefficients $\{F_k,\ k = 1 \ldots N\}$ to be estimated reliably, this finite crosstalk matrix must be nonsingular and reasonably well conditioned. This will be the case if $\mathbf{B}_N$ is approximately diagonal and none of the diagonal elements is close to zero. Off-diagonal elements do not necessarily make $\mathbf{B}_N$ singular, but they do decrease the determinant (for constant diagonal elements) and lead to noise amplification when an inversion of $\mathbf{B}_N$ is performed.[25]

On this basis BG suggested the following design strategy:

1. Choose the number of Fourier coefficients to be estimated (i.e., choose the spatial resolution).
2. Choose the system geometry in such a way as to minimize the off-diagonal elements of $\mathbf{B}_N$ and maximize the diagonal elements.
3. Check that $\beta_{kk'}$ is negligible for $k \le N$ and $k' > N$.

Step 1 is necessary in practice because the desired spatial resolution of the system imposes constraints on sampling and other aspects of the system design. Step 3 is a requirement to ensure that Fourier coefficients outside the band of interest are not aliased into the data set.

Step 2 may, in fact, not be achievable; it is possible that variation of a system design parameter in such a way as to minimize off-diagonal elements of $\mathbf{B}_N$ will also reduce diagonal elements. In these cases we need a figure of merit to quantify the trade-off, and a principal objective of this paper is to provide such figures of merit. As we shall see from the examples in Subsection 2.E, however, there may be one set of design parameters that influences mainly diagonal elements and another that influences mainly off-diagonal elements. In these happy circumstances, step 2 of the BG strategy is unambiguous.

E. Examples

To make the crosstalk concept more concrete, we now present several examples of its use. For a more complicated example, see BG on cone–beam tomography.

Example 1: Ideal Detector Array

Consider first an imaging system in which a perfect image is sampled by a regular array of detectors of finite size. For simplicity, we work in one dimension and specify the detector sensitivity functions by rect functions:

$$h_m(x) = \mathrm{rect}\left(\frac{x - m\Delta}{w}\right), \tag{14}$$

where $\mathrm{rect}(t) \equiv 1$ if $|t| < 1/2$ and zero otherwise. In this problem, Δ is the center-to-center spacing of the detectors and $w(\le\Delta)$ is the width of a single detector. We assume that $-M \le m \le M$, for a total of $2M + 1$ detectors, and that the support function $S(x) = \mathrm{rect}(x/L)$ with $L = (2M + 1)\Delta$ so that the detector array fits within the object support.

From the 1D counterpart of Eq. (7) we have

$$\Psi_{mk} = \int_{m\Delta - w/2}^{m\Delta + w/2} \exp(2\pi i \xi_k x)\mathrm{d}x = \exp(2\pi i \xi_k m\Delta) w \ \mathrm{sinc}(w\xi_k)\,, \tag{15}$$

where $\mathrm{sinc}(u) \equiv \sin(\pi u)/(\pi u)$. From Eq. (9) it follows that the diagonal elements of the crosstalk matrix are

$$\beta_{kk} = (2M + 1)w^2 \ \mathrm{sinc}^2(w\xi_k) \tag{16}$$

and that the off-diagonal elements are

$$\beta_{kk'} = w^2 \ \mathrm{sinc}(w\xi_k)\mathrm{sinc}(w\xi_{k'})\frac{\sin[(2M + 1)a]}{\sin(a)}\,, \tag{17}$$

where $a = \pi\Delta(\xi_k - \xi_{k'})$.

This simple example supports the interpretation of the crosstalk matrix in terms of transfer function and aliasing. The diagonal element β_{kk} is the square of the detector transfer function, evaluated at the spatial frequency ξ_k. The off-diagonal element $\beta_{kk'}$ is approximately zero if the Nyquist condition is satisfied for both components, i.e., both ξ_k and $\xi_{k'}$ are less than $1/2\Delta$.

The design parameters for this system are the detector spacing Δ and the detector width w. By inspection of Eqs. (16) and (17) we see that Δ affects the off-diagonal elements but not the diagonal ones. The detector width w affects all elements in principle, but note from Eq. (13) that it drops out of $\theta_{kk'}$, so the degree of orthogonality of different Fourier components as reflected in the data is independent of w. Thus we can separately manipulate the diagonal and the off-diagonal elements in the system design in this case.

The BG design strategy for this problem is elementary. After choosing spatial resolution in step 1, we minimize the off-diagonal elements of the finite crosstalk matrix by choosing Δ to satisfy the Nyquist condition for the maximum frequency, and we maximize the diagonal elements by choosing $w = \Delta$. Since we do not assume that the object is band limited, we must also invoke step 3 and check whether higher frequencies present in the objects of interest can alias significantly into the band chosen in step 1. If they cannot, the design is satisfactory.

If there is still significant aliasing (or if we are not sure how much is significant), however, the BG strategy provides us with no guidance. We could go back to step 1 and choose a higher maximum frequency, which would then lead us to smaller Δ and w and more detectors, but it is not obvious that this approach is advantageous since both diagonal and off-diagonal elements would thereby be reduced. It is precisely at this point that noise must be considered, and more detailed figures of merit, to be derived below, are needed.

Example 2: Incoherent Imaging

Next consider a 1D incoherent optical system with a rectangular pupil and a discrete detector array. The lens contributes a point-spread function $p_{\mathrm{lens}}(x)$, determined in general by both diffraction and aberrations, and there is additional blur from the finite detector size. The specification of the detector array and the object support are exactly as in Example 1. The sensitivity function is

$$h_m(x) = \int_{m\Delta - w/2}^{m\Delta + w/2} \mathrm{d}x' p_{\mathrm{lens}}(x' - x)\,. \tag{18}$$

When this form is substituted into the 1D version of Eq. (3), the familiar result is obtained that the mean output of the mth detector is the object convolved with p_{lens} and then integrated over the detector response function.

From Eq. (7), we obtain Ψ_{mk} by Fourier transforming the product of $h_m(x)$ and the support function:

$$\begin{aligned}\Psi_{mk} &= \int_{-L/2}^{L/2} \mathrm{d}x \int_{m\Delta - w/2}^{m\Delta + w/2} \mathrm{d}x' p_{\mathrm{lens}}(x' - x)\exp(2\pi i \xi_k x) \\ &= \int_{-L/2}^{L/2} \mathrm{d}x p_{\mathrm{tot}}(x - m\Delta)\exp(2\pi i \xi_k x)\,, \end{aligned} \tag{19}$$

where $p_{\mathrm{tot}}(x)$ is the total PSF, the convolution of $p_{\mathrm{lens}}(x)$, and the detector response function $\mathrm{rect}(x/w)$.

A useful approximation at this point is to assume that $p_{\mathrm{tot}}(x)$ is compact in comparison with the support function, so that L can be replaced by ∞. With this approximation we have

$$\Psi_{mk} \simeq P_{\mathrm{tot}}(\xi_k)\exp(2\pi i \xi_k m\Delta)\,, \tag{20}$$

where $P_{\mathrm{tot}}(\xi)$ is the 1D Fourier transform of $p_{\mathrm{tot}}(x)$. Equations (16) and (17) now hold if we simply replace $w\,\mathrm{sinc}(w\xi_k)$ with $P_{\mathrm{tot}}(\xi_k)$. A diagonal element β_{kk} depends on detector size, aperture size, wavelength, focal length, and any other lens parameters that influence P_{tot}. Again the overall weighting factor ${P_{\mathrm{tot}}}^*(\xi_k)P_{\mathrm{tot}}(\xi_{k'})$ affects the off-diagonal element $\beta_{kk'}$ but drops out of $\theta_{kk'}$. The lengths of the vectors $\mathcal{H}\Phi_k$ and $\mathcal{H}\Phi_{k'}$ depend on w, and the angle between them depends on Δ.

Example 3: Shift-Invariant Imaging Systems

In Example 2 we made a step in the direction of modeling the imaging system as a continuous, linear, shift-invariant (LSIV) imaging system when we assumed that the object support was large compared with the support of $p_{\mathrm{tot}}(x)$. To get to a true continuous LSIV model we must also assume that the sample spacing Δ is small. Then we can treat the sum over m in Eq. (9) as an integral. Retaining the approximation that L is large, we have

$$\begin{aligned}\beta_{kk'} &\simeq {P_{\mathrm{tot}}}^*(\xi_k)P_{\mathrm{tot}}(\xi_{k'}) \sum_{m=-M}^{M} \exp[2\pi i(\xi_{k'} - \xi_k)m\Delta] \\ &\simeq {P_{\mathrm{tot}}}^*(\xi_k)P_{\mathrm{tot}}(\xi_{k'}) \frac{1}{\Delta} \int_{-L/2}^{L/2} \mathrm{d}x \exp[2\pi i(\xi_{k'} - \xi_k)x]\,. \end{aligned} \tag{21}$$

Recall that $L = (2M + 1)\Delta$, so large L implies large M. For $L \to \infty$, the integral vanishes unless $k = k'$. Thus the crosstalk matrix in this case is automatically diagonal, and we have

$$\beta_{kk'} \simeq \frac{L}{\Delta}|P_{\mathrm{tot}}(\xi_k)|^2 \delta_{kk'}\,. \tag{22}$$

The corresponding result in two or three dimensions follows analogously, with L being replaced by the area A or the volume V of the region of support and Δ by Δ^2 or Δ^3 as appropriate. Thus a continuous LSIV system has a diagonal crosstalk matrix, and the diagonal element is proportional to $|P_{\mathrm{tot}}(\xi_k)|^2$, justifying our designation of β_{kk} as a generalized transfer function.

Example 4: Parallel-Beam Tomography

Consider, as a slightly more complicated example of the crosstalk matrix, a 2D tomographic problem in which a function $f(\mathbf{r})$ supported on a circle of radius R is sampled with a set of thin rays of finite width w. The orientation of a ray is specified by its distance p from the origin and its angle ϕ with respect to the x axis; we obtain a tomographic data set by stepping p and ϕ through discrete values. The detector index m must now include a specification of both p and ϕ, and it is convenient to replace the single index m by two indices m and n. The sample values of p are taken as $p_m = mR/M_p$, $-M_p \le m \le M_p$, and the values of ϕ are $\phi_n = n\pi/M_\phi$, $0 \le n \le M_\phi - 1$. The total number of measurements is thus $(2M_p + 1)M_\phi$, and all sums over m in the formalism above are to be replaced by double sums over m and n for this example. With these conventions the sensitivity function is given by

$$h_{mn}(\mathbf{r}) = \frac{1}{w}\ \mathrm{rect}\left[\frac{p_m - r\cos(\theta - \phi_n)}{w}\right], \tag{23}$$

where the polar coordinates of the 2D vector $\mathbf{r}$ are (r, θ), θ being measured from the x axis. Readers familiar with tomography[29] will recognize $h_{mn}(\mathbf{r})$ as a finite-width, sampled version of the line delta function $\delta[p - r\cos(\theta - \phi)]$.

If we assume that the ray is thin compared with the object support, $w \ll R$, the integral in Eq. (7) is easy to perform, with the result (see BG for details) that

$$\begin{aligned}\Psi_{mnk} = 2\sqrt{R^2 - {p_m}^2}\,&\mathrm{sinc}[w\rho_k \cos(\phi_n - \alpha_k)] \\ &\times \mathrm{sinc}[2\sqrt{R^2 - {p_m}^2}\,\rho_k \sin(\phi_n - \alpha_k)] \\ &\times \exp\{2\pi i[\rho_k p_m \cos(\phi_n - \alpha_k)]\}\,, \end{aligned} \tag{24}$$

where ρ_k and α_k are the polar coordinates of $\boldsymbol{\rho}_k$, with α_k being measured from the x axis. The diagonal elements of the crosstalk matrix are then given by

$$\begin{aligned}\beta_{kk} = \sum_{m,n} 4(R^2 - {p_m}^2)&\mathrm{sinc}^2[w\rho_k \cos(\phi_n - \alpha_k)] \\ &\times \mathrm{sinc}^2[2\sqrt{R^2 - {p_m}^2}\,\rho_k \sin(\phi_n - \alpha_k)], \end{aligned} \tag{25}$$

and the off-diagonal elements are given by a somewhat more complicated expression obtainable by substitution of Eq. (24) into Eq. (9).

The main qualitative point to note is that the product of sinc functions in Eq. (24) defines a long, thin region of the spatial-frequency plane, and a particular measurement m, n provides significant information about Fourier coefficient F_k only if $\boldsymbol{\rho}_k$ lies in this region. We therefore refer to the region as the contribution region for measurement m, n. The contribution region is centered on the origin in frequency space, and its orientation is perpendicular to the ray defined by m and n. The width of the contribution region in the perpendicular direction is $1/w$. The width of the region in the direction parallel to the ray is given by $1/[2(R^2 - {p_m}^2)^{1/2}]$, so it depends on the particular ray chosen, but it will be small compared with the perpendicular width for almost all rays. Conventional treatments of tomography often ignore the finite width of the rays and the finite size of the object support, in which case the contribution region becomes a line delta function through the origin in frequency space. We can recover this limit formally by letting $w \to 0$ and $R \to \infty$.

The system design parameters in this example are w, M_p, and M_ϕ. Although all three of these parameters enter into Eq. (25), M_p affects β_{kk} only weakly. The stronger effects on the diagonal elements are the ray width w and the number of angular samples M_ϕ, both of which can strongly reduce β_{kk} for large ρ_k. If ρ_k is smaller than $1/w$ and M_ϕ is large, it can be shown that β_{kk} is proportional to $1/\rho_k$, as expected from conventional tomographic theory.[29] When ρ_k becomes larger than $1/w$, there is an additional attenuation of the transfer function from the first sinc^2 in Eq. (25), so $1/w$ is an approximate frequency cutoff for the system.

The effect of M_ϕ is more subtle. If we have a small number of angular samples, there can be some frequencies $\boldsymbol{\rho}_k$ with magnitude less than $1/w$ that fall into the gaps between contribution regions for different n, and such components will have a small β_{kk}. Thus *angular* sampling can affect diagonal elements. Sampling in p, on the other hand, affects mainly off-diagonal elements. Two frequency vectors $\boldsymbol{\rho}_k$ and $\boldsymbol{\rho}_{k'}$ can be aliased if they are approximately parallel and the Nyquist condition is not satisfied for the one with the larger magnitude, so $\beta_{kk'}$ is strongly influenced by M_p.

More details on the crosstalk matrix for this example, as well as a numerical study to make these qualitative observations more precise, are the subject of ongoing research. The main point for present purposes, however, is that we again have a situation where some design parameters (w and M_ϕ) affect mainly diagonal elements, while another design parameter (M_p) is more important for off-diagonal elements.

3. NOISE MODELS AND FISHER INFORMATION

A. Definition and Properties of the Fisher Information Matrix

Given an $M \times 1$ random data vector $\mathbf{g}$ and an $N \times 1$ vector $\boldsymbol{\theta}$ of unknown (but nonrandom) real parameters, the Fisher information matrix $\mathbf{J}_\theta$ is defined in terms of its elements $[J_\theta]_{kk'}$ in two mathematically equivalent ways[26–28]:

$$[J_\theta]_{kk'} = \left\langle \frac{\partial}{\partial \theta_{k'}} \ln p(\mathbf{g}|\boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} \ln p(\mathbf{g}|\boldsymbol{\theta}) \right\rangle = -\left\langle \frac{\partial^2}{\partial \theta_{k'} \partial \theta_k} \ln p(\mathbf{g}|\boldsymbol{\theta}) \right\rangle, \tag{26}$$

where $p(\mathbf{g}|\boldsymbol{\theta})$ is the conditional probability-density function of the data (or the probability if the data are discrete valued) and $\langle \ldots \rangle$ denotes an expectation with respect to this density. It can be shown that $\mathbf{J}_\theta$ is positive semidefinite (positive definite if nonsingular).

Let $\hat{\boldsymbol{\theta}}$ be any unbiased estimator of $\boldsymbol{\theta}$ and assume that $\mathbf{J}_\theta$ is nonsingular. The Cramer–Rao (CR) inequality[26–28] says that

$$\mathrm{var}(\hat{\boldsymbol{\theta}}_k) \geq [\mathbf{J}_{\boldsymbol{\theta}}{}^{-1}]_{kk} \,. \tag{27}$$

An estimator that achieves this lower bound is called *efficient.*

A related inequality places a bound on the covariance matrix of the estimator. As in paper I, we denote this matrix by $\hat{\mathbf{K}}_{\boldsymbol{\theta}}$ in order to avoid ornaments on subscripts, but the caret over $\mathbf{K}$ implies that it is the covariance matrix of the estimate, not an estimate of the covariance matrix. This matrix must satisfy

$$\hat{\mathbf{K}}_{\boldsymbol{\theta}} \geq \mathbf{J}_{\boldsymbol{\theta}}{}^{-1} \,. \tag{28}$$

This inequality must be interpreted carefully. In general, the statement $\mathbf{A} \geq \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ matrices, means only that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. If $\mathbf{A}$ and $\mathbf{B}$ are themselves positive definite, however, it also follows that $\mathrm{tr}(\mathbf{AC}) \geq \mathrm{tr}(\mathbf{BC})$, where tr denotes trace and $\mathbf{C}$ is any positive-definite $N \times N$ matrix, and that $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$ in the same sense.[30]

The inequality in relation (28) becomes an equality if $\hat{\boldsymbol{\theta}}$ is efficient and unbiased; then relation (28) can be legitimately interpreted as an element-by-element equality, and the covariance matrix of the estimates is the inverse of the Fisher information matrix.[27,28]

One other bound will turn out to be important in what follows. For any positive-definite $N \times N$ matrix $\mathbf{A}$, it is known that $A_{nn} \geq 1/[\mathbf{A}^{-1}]_{nn}$, with equality if and only if $\mathbf{A}$ is diagonal. (A derivation of this inequality is given in Appendix A.) Thus inequality (27) generalizes to

$$\mathrm{var}(\hat{\theta}_k) \geq [\mathbf{J}_{\boldsymbol{\theta}}{}^{-1}]_{kk} \geq \frac{1}{[\mathbf{J}_{\boldsymbol{\theta}}]_{kk}} \,, \tag{29}$$

where the first inequality is an equality for an efficient estimator and the second is an equality for a diagonal Fisher matrix. The latter form is particularly useful when the number of unknown parameters is large so that inversion of $\mathbf{J}_{\boldsymbol{\theta}}$ is not feasible.

Although the second inequality in relation (29) is sometimes mentioned in books on estimation theory, it is usually dismissed as a weak bound that will give little information on the usefulness of the estimator. When we are evaluating the measurement system rather than the estimator, however, the second inequality is the *stronger* one. For the estimator, we want a greatest lower bound so that we can evaluate how close to optimum the estimator is. Given the system, an optimum estimator is one for which the variance is as close as possible to the CR bound, so it is advantageous to know the greatest lower bound. If we presume the existence of an efficient estimator, on the other hand, an optimum system is one for which the CR bound is *smallest*, and the second inequality tells us how small it can be.

B. Application to Estimation of Fourier Coefficients

There are two difficulties in applying the results of Subsection 3.A to the imaging problem as formulated in BG, where the parameters to be estimated are the Fourier coefficients $\{F_k\}$. The first is that the parameters are complex. We could, of course, reformulate the problem as the estimation of strictly real parameters, but the complex notation is more compact. We therefore define the Fisher information matrix for the $\{F_k\}$ as

$$J_{kk'} = -\left\langle \frac{\partial^2}{\partial F_k{}^* \partial F_{k'}} \ln p(\mathbf{g}|\mathbf{F}) \right\rangle, \tag{30}$$

where $p(\mathbf{g}|\mathbf{F})$ is the conditional probability-density function of the data, given a specific object. Since the object is fully specified by the vector $\mathbf{F}$ of Fourier coefficients, $p(\mathbf{g}|f(\mathbf{r})) = p(\mathbf{g}|\mathbf{F})$. Since the object is fixed, the only randomness comes from the noise vector $\boldsymbol{\epsilon}$ in Eq. (4), and the expectation is thus over that random variable. It is easy to show that $\mathbf{J}$ defined this way is Hermitian and positive semidefinite.

The second problem is that there are an infinite number of Fourier coefficients. In paper I we noted that not all parameters that one might associate with an object are estimable. A parameter is termed estimable if it is possible to find an unbiased estimator of it for all true values of the parameter. With this definition, any Fourier component that does not contribute to the data is not estimable, since any estimate of it formed from the data would necessarily be independent of the true value. In terms of the crosstalk matrix, F_k is not estimable if $\beta_{kk} = 0$. This problem needs to be taken into account in the choice of the number of Fourier coefficients to be estimated. This number, denoted N, must be such that $\beta_{kk} \neq 0$ for $1 \leq k \leq N$. If this condition is satisfied, we simply use Eq. (30) as the definition of the finite $(N \times N)$ Fisher information matrix, denoted $\mathbf{J}_N$. If $\mathbf{J}_N$ is nonsingular and no higher frequencies are aliased into the band of interest (step 3 of the BG design strategy), all the F_k for $k \leq N$ are estimable.

C. Noise Models

To evaluate Eq. (30) we need a specific model for the probability-density function for $\boldsymbol{\epsilon}$. If we assume Gaussian noise, the most general form for $p(\mathbf{g}|\mathbf{F})$ is a multivariate Gaussian with covariance matrix $\mathbf{K}_\epsilon$ and mean vector $\overline{\mathbf{g}} = \boldsymbol{\Psi}\mathbf{F}$. We thus have

$$p(\mathbf{g}|\mathbf{F}) = \frac{1}{(2\pi)^{M/2}\sqrt{\det(\mathbf{K}_\epsilon)}} \times \exp\left[-\frac{1}{2}(\mathbf{g}-\overline{\mathbf{g}})^t \mathbf{K}_\epsilon^{-1}(\mathbf{g}-\overline{\mathbf{g}})\right], \quad (31)$$

where $\det(\mathbf{K}_\epsilon)$ is the determinant of the noise covariance matrix and the superscript t denotes transpose. Differentiation of Eq. (31) yields the well-known result

$$J_{kk'} = \sum_{m=1}^{M}\sum_{m'=1}^{M} \Psi_{mk}^*(\mathbf{K}_\epsilon^{-1})_{mm'}\Psi_{m'k'} = [\boldsymbol{\Psi}^\dagger \mathbf{K}_\epsilon^{-1}\boldsymbol{\Psi}]_{kk'}. \quad (32)$$

In most digital imaging situations with discrete detectors, the raw measurements are statistically independent (for a nonrandom object), so

$$[\mathbf{K}_\epsilon]_{mm'} = \sigma_m^2 \delta_{mm'}, \quad (33)$$

where σ_m^2 is the variance of the mth measurement and $\delta_{mm'}$ is the Kronecker delta. With this noise covariance, we have

$$J_{kk'} = \sum_{m=1}^{M} \frac{\Psi_{mk}^*\Psi_{mk'}}{\sigma_m^2}. \quad (34)$$

This expression looks much like the crosstalk matrix [cf. Eq. (9)], but there is a weighting factor $1/\sigma_m^2$ inside the sum. If all variances are the same (often a good model for signal-independent electronic noise), then $\mathbf{J} = \mathbf{B}/\sigma^2$.

In many photon-limited imaging situations, the probability law on the data is Poisson, and we have

$$p(\mathbf{g}|\mathbf{F}) = \prod_{m=1}^{M} \frac{(\overline{g}_m)^{g_m}}{g_m!}\exp(-\overline{g}_m) = \prod_{m=1}^{M} \frac{(\boldsymbol{\Psi}\mathbf{F})_m{}^{g_m}}{g_m!}\exp[-(\boldsymbol{\Psi}\mathbf{F})_m]. \quad (35)$$

Differentiation of the second form of Eq. (35) and use of Eq. (30) lead to

$$J_{kk'} = \sum_{m=1}^{M} \frac{\Psi_{mk}^*\Psi_{mk'}}{(\boldsymbol{\Psi}\mathbf{F})_m} = \sum_{m=1}^{M} \frac{\Psi_{mk}^*\Psi_{mk'}}{\overline{g}_m}. \quad (36)$$

Since detectors for which $\overline{g}_m = 0$ necessarily receive no photons, they contribute a constant factor of unity to Eq. (35); after differentiation they disappear, so terms with $\overline{g}_m = 0$ are to be deleted from the sum in Eq. (36). Equation (36) was also obtained recently by Hero and Fessler.[31]

A comparison of Eqs. (32) and (36) shows that the Poisson and multivariate normal model yield the same form for $\mathbf{J}$ except that, in the Poisson case,

$$[\mathbf{K}_\epsilon]_{mm'} = (\boldsymbol{\Psi}\mathbf{F})_m \delta_{mm'}. \quad (37)$$

Note, however, that the noise covariance matrix, and hence the Fisher information matrix, depends on the object in this case.

Again, the Fisher information resembles the crosstalk matrix except for a weighting factor in the sum. If the data have low contrast, such that $(\boldsymbol{\Psi}\mathbf{F})_m$ is approximately a constant independent of m, then we again have $\mathbf{J} \simeq \mathbf{B}/\sigma^2$, but this approximation is often not valid with Poisson noise.

4. BOUNDS ON VARIANCE OF ESTIMATES OF FOURIER COEFFICIENTS

A. Fundamentals of the Cramer–Rao Bound

For any unbiased estimate $\hat{F}_k$ of the Fourier coefficient F_k, the variance is defined as

$$\mathrm{var}(\hat{F}_k) \equiv \langle|\hat{F}_k - F_k|^2\rangle, \quad (38)$$

where the expectation is over realizations of $\boldsymbol{\epsilon}$ for fixed $\mathbf{F}$. We would like to use relation (27) or (29) to set a lower bound on $\mathrm{var}(\hat{F}_k)$, but we cannot use the infinite Fisher information matrix for this purpose because it is necessarily singular; since $\mathbf{J} = \boldsymbol{\Psi}^\dagger\mathbf{K}_\epsilon^{-1}\boldsymbol{\Psi}$, the rank of $\mathbf{J}$ cannot exceed the number of measurements M. If, however, N coefficients are estimable and $N \leq \mathrm{rank}(\mathbf{J})$, only the finite Fisher information matrix $\mathbf{J}_N$ is relevant, and we have, from relation (29),

$$\mathrm{var}(\hat{F}_k) \geq (\mathbf{J}_N^{-1})_{kk} \geq \frac{1}{J_{kk}}, \qquad k = 1 \ldots N. \quad (39)$$

The first inequality becomes an equality for an efficient estimator, and the second becomes an equality for a diagonal Fisher information matrix. In other words, as discussed in Subsection 3.A, the first inequality relates to the image-reconstruction process, while the second relates to system design.

Of course, the considerations in the preceding paragraph presume the existence of an efficient, unbiased estimator of the Fourier coefficients. It is shown in Appendix B that such an estimator exists for the multivariate Gaussian noise model and that a good approximation to one exists for Poisson noise.

B. Fisher Information and Cross Talk for Poisson Data

From Eq. (36) the diagonal element J_{kk} on the right-hand side of relation (39) is given for Poisson data by

$$J_{kk} = \sum_{m=1}^{M} \frac{|\Psi_{mk}|^2}{\overline{g}_m} . \tag{40}$$

Since $|\Psi_{mk}|^2$ varies quadratically with the exposure time τ while $\overline{g}_m$ varies linearly, Eq. (40) shows that J_{kk} varies as τ and the variance bound as $1/\tau$ for Poisson noise.

As we have seen, one way to relate the Fisher and crosstalk matrices is to assume that all the data variances are equal, which in the Poisson case means that the data have very low contrast. In Eq. (40), if all of the mean data values $\overline{g}_m$ are equal to a constant $\overline{g}$, J_{kk} becomes $\beta_{kk}/\overline{g}$.

This extreme low-contrast approximation is unlikely to be valid, but a related approximation is more defensible. To introduce it we treat $|\Psi_{mk}|^2$, when properly normalized, as a probability and use it to define a mean effective number of counts,

$$\overline{g}_{\text{eff}}(k) \equiv \sum_{m=1}^{M} |\Psi_{mk}|^2 \overline{g}_m \bigg/ \sum_{m=1}^{M} |\Psi_{mk}|^2 . \tag{41}$$

This definition weights each mean measurement $\overline{g}_m$ by $|\Psi_{mk}|^2$; this weighting factor is interpreted as the relative strength of the contribution of the kth Fourier component to the mth measurement. Thus $\overline{g}_{\text{eff}}(k)$ is the weighted average number of photons that are useful for estimation of F_k. If $|\Psi_{mk}|^2$ is independent of the detector index m, as it is for an LSIV system [see Eq. (20)], all photons are equally useful and $\overline{g}_{\text{eff}}(k)$ is the mean total number of counts divided by the number of detectors. For tomographic systems, however, only a subset of the detectors (the critical detectors in BG) is useful for estimation of a particular F_k, and $\overline{g}_{\text{eff}}(k)$ is the weighted average number of photons measured by these detectors.

The actual mean number $\overline{g}_m$ can be written in terms of deviations from this effective mean as

$$\overline{g}_m = \overline{g}_{\text{eff}}(k) + \delta\overline{g}_m(k) , \tag{42}$$

and a Taylor series yields

$$\frac{1}{\overline{g}_m} = \frac{1}{\overline{g}_{\text{eff}}(k)} - \frac{\delta\overline{g}_m(k)}{[\overline{g}_{\text{eff}}(k)]^2} + \frac{[\delta\overline{g}_m(k)]^2}{[\overline{g}_{\text{eff}}(k)]^3} + \dots . \tag{43}$$

Note that the center of expansion depends on the spatial-frequency index k. Substituting the Taylor expansion into Eq. (40), we find that

$$J_{kk} = \frac{\sum_{m=1}^{M} |\Psi_{mk}|^2}{\overline{g}_{\text{eff}}(k)} + \frac{\sum_{m=1}^{M} |\Psi_{mk}|^2 [\delta\overline{g}_m]^2}{[\overline{g}_{\text{eff}}(k)]^3} + \dots . \tag{44}$$

The first term is $\beta_{kk}/\overline{g}_{\text{eff}}(k)$, and the second is a positive-semidefinite correction term. We shall refer to the approximation of neglecting all but the leading term in Eq. (44) as the moderate-contrast Poisson model.

The moderate-contrast model is valid if the variations in $\overline{g}_m$ are small, but this condition is too restrictive. It is sufficient that variations in $\overline{g}_m$ be small over the subset of detectors for which the factor $|\Psi_{mk}|^2$ is appreciable. Small variations over this subset may be more likely than small variations when all detectors are considered. For example, in a tomographic system, $\overline{g}_m$ will be small for rays that pass through only a small portion of the object, but $|\Psi_{mk}|^2$ is also small for rays that pass through a small portion of the region of support (see BG). These two effects are compensating, at least for objects that substantially fill the region of support, and J_{kk} can be well approximated by the leading term in Eq. (44) even though $\overline{g}_m$ goes to zero for some detectors.

We have performed a few numerical studies of tomographic systems to check the validity of the moderate-contrast model. These studies are beyond the scope of this paper because they refer to specific imaging systems, but they show the approximation to be rather robust. Note that contrast here refers to variations in mean number of counts in the data set, not to object contrast. In tomographic problems, high-contrast objects can easily lead to low-contrast data.

On the moderate-contrast Poisson model we have

$$\text{var}(\hat{F}_k) \geq \overline{g}_{\text{eff}}(k)/\beta_{kk} . \tag{45}$$

The key difference between this bound and the one obtained on the extreme low-contrast assumption is that the numerator is now a function of spatial frequency $\boldsymbol{\rho}_k$.

5. CLASSIFICATION TASKS

We turn next to specification of image quality by observer performance on classification tasks. The observer considered is either the ideal Bayesian observer or the ideal linear (Hotelling) observer.[13,17,32] The ideal observer sets an upper limit to the performance of any observer. The Hotelling observer is the optimum observer among those constrained to perform only linear operations on the data. For a binary (two-alternative) classification task, the Hotelling observer is essentially the Fisher linear discriminant except that the observer is presumed to have knowledge of ensemble means and covariance matrices instead of having to estimate them from the data. The figure of merit for this observer, sometimes called the Hotelling trace, is proportional to the ensemble Mahalanobis distance. An important advantage of the Hotelling observer is that it is analytically tractable in many cases in which the ideal observer is not. Moreover, the Hotelling observer or a simple modification of it has been found to predict human performance in many cases.[13,17,21,22,32–37]

We presume initially that the observer has access to the raw data and uses those data to make a decision, but later we consider the case in which the observer has access only to the estimates of the Fourier coefficients. The treatment relies heavily on paper I, but the object model is continuous here.

A. SKE/BKE Detection

The simplest binary classification task is one where it is known that the object is either exactly $f_1(\mathbf{r})$ or exactly $f_2(\mathbf{r})$. If we think of $f_1(\mathbf{r})$ as a background and $f_2(\mathbf{r})$ as the same background with some known signal superimposed, this classification task becomes a signal-detection task. We use the language of detection in what follows, but the results are equally applicable to discrimination between two exactly specified objects.

For this task and either Gaussian or Poisson noise, the ideal Bayesian observer given access to $\mathbf{g}$ would compute a linear discriminant function of the form $\lambda(\mathbf{g}) = \mathbf{w}^t\mathbf{g}$. The form of the template $\mathbf{w}$ depends on the noise model assumed. For multivariate normal noise, the template is[38]

$$\mathbf{w} = \mathbf{K}_\epsilon{}^{-1}\Delta\overline{\mathbf{g}}, \tag{46}$$

where $\Delta\overline{\mathbf{g}}$ is the mean difference signal given, for a continuous-to-discrete mapping, by

$$\Delta\overline{\mathbf{g}} = \mathcal{H}\{f_2(\mathbf{r}) - f_1(\mathbf{r})\} = \mathbf{\Psi}[\mathbf{F}_2 - \mathbf{F}_1] \equiv \mathbf{\Psi}\Delta\mathbf{F}. \tag{47}$$

For Poisson noise, the form of the ideal-observer template is slightly different,[19] but Eq. (46) is a good approximation if $\mathbf{K}_\epsilon$ is the appropriate diagonal matrix and $\overline{g}_m$ is not too small. In any event, the linear discriminant function specified in Eq. (46) is the template used by the Hotelling observer on this problem, even with low-count Poisson noise. In the Gaussian SKE/BKE case, the ideal and Hotelling observers are identical.

The performance of any observer on a binary classification task is specified by a signal-to-noise ratio (SNR) defined as the difference in means of $\lambda(\mathbf{g})$ under the two hypotheses divided by the square root of the average variance. For the problem at hand, the SNR for the Hotelling observer is discussed in Refs. 13, 17, 18, and 32. The result is

$$[\mathrm{SNR(SKE/BKE, Hot, }\mathbf{g})]^2 = \Delta\overline{\mathbf{g}}^\dagger\mathbf{K}_\epsilon{}^{-1}\Delta\overline{\mathbf{g}} = \sum_{m=1}^{M}\frac{[\Delta\overline{g}_m]^2}{\sigma_m{}^2}, \tag{48}$$

where the last form assumes statistically independent but not necessarily identically distributed Gaussian or Poisson noise. (Hot stands for Hotelling.) The SNR notation, though cumbersome, indicates the task, the observer, and the data vector given to the observer.

We wish to rewrite this SNR in terms of the Fisher information matrix appropriate to estimation of the Fourier coefficients. From Eqs. (47) and (48) we have, for statistically independent noise,

$$\begin{aligned}[\mathrm{SNR(SKE/BKE, Hot, }\mathbf{g})]^2 &= \sum_{m=1}^{M}\frac{[(\mathbf{\Psi}\Delta\mathbf{F})_m]^2}{\sigma_m{}^2}\\ &= \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\Delta F_k{}^*\Delta F_{k'}\sum_{m=1}^{M}\frac{\Psi_{mk}{}^*\Psi_{mk'}}{\sigma_m{}^2} = \Delta\mathbf{F}^\dagger\mathbf{J}\Delta\mathbf{F}. \end{aligned}\tag{49}$$

For identically distributed noise, such as Poisson noise in the extreme low-contrast limit, $\sigma_m{}^2 = \sigma^2$ for all m, and the SNR expression simplifies further to

$$\begin{aligned}&[\mathrm{SNR(SKE/BKE, Hot, }\mathbf{g})]^2\\ &= \frac{1}{\sigma^2}\sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\Delta F_k{}^*\Delta F_{k'}\sum_{m=1}^{M}\Psi_{mk}{}^*\Psi_{mk'} = \frac{1}{\sigma^2}\Delta\mathbf{F}^\dagger\mathbf{B}\Delta\mathbf{F}.\end{aligned}\tag{50}$$

In both Eq. (49) and Eq. (50), the sums over k and k' are infinite. (Actually, each runs over the infinite 3D lattice of points in frequency space as discussed in Subsection 2.A.) Since the observer has access to the entire data vector $\mathbf{g}$, any Fourier component of $\Delta\mathbf{F}$ that contributes to the data affects the SNR. Thus the infinite Fisher information matrix enters into Eq. (49), and the infinite crosstalk matrix enters into Eq. (50), and it does not matter that both are singular.

If $\mathbf{B}$ is diagonal, Eq. (50) takes the simple form

$$[\mathrm{SNR(SKE/BKE, Hot, }\mathbf{g})]^2 = \frac{1}{\sigma^2}\sum_{k=1}^{\infty}\beta_{kk}|\Delta F_k|^2. \tag{51}$$

This result makes sense when we recall the interpretation of β_{kk} as a transfer function describing how strongly a given Fourier component contributes to the data; in this case it is the Fourier component of the difference signal that is important.

B. Ideal/Hotelling Observer Operating on a Reconstructed Image

Consider an SKE/BKE detection problem in which the data set presented to the observer is not the raw data $\mathbf{g}$ but a reconstructed image. If the reconstruction algorithm yields unbiased, efficient estimates of the first N Fourier coefficients, we can equivalently say that the data presented to the observer are $\{\hat{F}_k,\ k = 1 \ldots N\}$. The SNR for the Hotelling observer in this case is given by an expression like the first form in Eq. (48) but with $\hat{\mathbf{F}}$ as the data in place of $\mathbf{g}$:

$$[\mathrm{SNR(SKE/BKE, Hot, }\hat{\mathbf{F}})]^2 = \Delta\overline{\hat{\mathbf{F}}}^\dagger\hat{\mathbf{K}}_{\mathbf{F}}{}^{-1}\Delta\overline{\hat{\mathbf{F}}}, \tag{52}$$

where $\Delta\overline{\hat{\mathbf{F}}}$ is the mean of the estimate of $\Delta\mathbf{F}$. Because the estimate is assumed to be unbiased, however, $\Delta\overline{\hat{\mathbf{F}}} = \Delta\mathbf{F}$. Also, if we assume an efficient estimate, $\mathbf{K}_{\hat{\mathbf{F}}}$ is just the inverse of the finite Fisher information matrix by Eq. (28). Thus the SNR looks exactly like the one in Eq. (49) but with the finite Fisher information matrix:

$$[\mathrm{SNR(SKE/BKE, Hot, }\hat{\mathbf{F}})]^2 = \Delta\mathbf{F}^\dagger\mathbf{J}_N\Delta\mathbf{F}. \tag{53}$$

Equation (53) shows how the choice of N affects performance on this task. If $|\Delta F_k|$ is small for k greater than some N_0, increasing N beyond N_0 does not gain much in performance.

C. Random Backgrounds

An advantage of the Hotelling figure of merit is that it can be calculated for random signals and backgrounds, something that is usually not possible for the Bayesian ideal observer. In essence, the Hotelling approach replaces the data covariance matrix with an intraclass scatter matrix $\mathbf{S}_2$ that accounts for both measurement noise and object variability (see paper I for details).

Consider first a Hotelling observer with access to the raw data $\mathbf{g}$ and the task of detection of an exactly known signal on a random background (RBG). The difference in mean data vectors is still given by Eq. (47), and the $\mathbf{S}_2$ matrix for this problem is

$$\mathbf{S}_{2\mathbf{g}} = \mathbf{\Psi}\mathbf{S}_{2\mathbf{F}}\mathbf{\Psi}^{\dagger} + \overline{\mathbf{K}}_{\boldsymbol{\epsilon}}\,, \tag{54}$$

where $\mathbf{S}_{2\mathbf{F}}$ is the intraclass scatter matrix for the object Fourier coefficients and the overbar on $\mathbf{K}_{\boldsymbol{\epsilon}}$, needed only in the case of signal-dependent (Poisson) noise, indicates an average over realizations of the random background (see paper I).

The figure of merit for this problem is thus

$$[\mathrm{SNR(SKE/RBG, Hot, }\mathbf{g})]^2 = [\mathbf{\Psi}\Delta\mathbf{F}]^{\dagger}[\mathbf{\Psi}\mathbf{S}_{2\mathbf{F}}\mathbf{\Psi}^{\dagger} + \overline{\mathbf{K}}_{\boldsymbol{\epsilon}}]^{-1}\mathbf{\Psi}\Delta\mathbf{F}\,. \tag{55}$$

A special case of this expression will be discussed below, but first we derive the corresponding result for a Hotelling observer given access only to $\hat{\mathbf{F}}$. If we assume that the estimates are unbiased and efficient, then the mean difference vector is $\Delta\mathbf{F}$. The appropriate $\mathbf{S}_2$ matrix is the one for $\hat{\mathbf{F}}$. We denote this matrix $\hat{\mathbf{S}}_{\mathbf{F}}$, where again the caret over $\mathbf{S}$ implies that it is an intraclass scatter matrix of the estimate, not an estimate of the intraclass scatter matrix. A derivation parallel to one given in paper I shows that

$$\hat{\mathbf{S}}_{2\mathbf{F}} = \mathbf{S}_{2\mathbf{F}} + \overline{\mathbf{J}}_N{}^{-1}\,, \tag{56}$$

where $\overline{\mathbf{J}}$ is the Fisher information matrix averaged over realizations of the background. This averaging has no effect with signal-independent noise, but with Poisson noise it has the effect of replacing $\mathbf{F}$ by its ensemble average in Eq. (36). Since only N Fourier coefficients are estimated, both $\hat{\mathbf{S}}_{2\mathbf{F}}$ and $\mathbf{S}_{2\mathbf{F}}$ must be interpreted as $N \times N$ matrices in Eq. (56). The SNR for detection in a random background is then

$$[\mathrm{SNR(SKE/RBG, Hot, }\hat{\mathbf{F}})]^2 = \Delta\mathbf{F}^{\dagger}[\mathbf{S}_{2\mathbf{F}} + \overline{\mathbf{J}}_N{}^{-1}]^{-1}\Delta\mathbf{F}\,. \tag{57}$$

D. Stationary Lumpy Backgrounds

To interpret Eqs. (55) and (57) we use a model in which the background is a stationary Gaussian random process. In its continuous form we have used this model, which we call the lumpy-background (LBG) paradigm, in a variety of theoretical and psychophysical studies.[20–22,33,36,37] In the present paper we use it for the first time in conjunction with a model of the imaging system as a continuous-to-discrete mapping. The mathematical statement of the LBG paradigm in this context is that the covariance matrix for the object Fourier coefficients is diagonal:

$$[\mathbf{K}_{\mathbf{F}}]_{kk'} \equiv \langle [F_k - \overline{F}_k]^*[F_{k'} - \overline{F}_{k'}]\rangle = W_k\delta_{kk'}\,, \tag{58}$$

where both the overbar and the angle brackets denote an average over the ensemble of random objects.

To see the implications of this form in the continuous-object domain, we compute the continuous autocorrelation function of $f(\mathbf{r})$, with the result that

$$\begin{aligned} R_f(\mathbf{r}, \mathbf{r}') &\equiv \langle [f(\mathbf{r}) - \overline{f}(\mathbf{r})][f(\mathbf{r}') - \overline{f}(\mathbf{r}')]\rangle \\ &= \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} W_k\delta_{kk'}\Phi_k{}^*(\mathbf{r})\Phi_{k'}(\mathbf{r}') \\ &= S(\mathbf{r})S(\mathbf{r}')\sum_{k=1}^{\infty} W_k \exp[2\pi i\boldsymbol{\rho}_k\cdot(\mathbf{r} - \mathbf{r}')]\,. \end{aligned} \tag{59}$$

Because of the two support functions, $R_f(\mathbf{r}, \mathbf{r}')$ is zero if either $\mathbf{r}$ or $\mathbf{r}'$ lies outside the known support of the object. The sum, however, is a function of only $\mathbf{r} - \mathbf{r}'$, so it represents a wide-sense stationary random process. We refer to this product of a stationary random process and a support function as a quasi-stationary random background. Equation (59) also shows that W_k can be interpreted as a power spectral density, in the sense that it is the coefficient in a Fourier *series* for the continuous autocorrelation function.

We emphasize that the background is stationary *in the ensemble-average sense* in this model. Any particular realization of the lumpy background might have one or a few blobs or other localized features that can interfere with the detection task, but there is no preferred location within the region of support for these features. Thus stationarity of the background is not as restrictive an assumption as it might appear.

Since we are considering the detection of an exactly known signal on a random background, the covariance matrix of $\mathbf{F}$ is the same whether or not a signal is present. Thus the intraclass scatter matrix $\mathbf{S}_{2\mathbf{F}}$ in Eq. (55) or Eq. (57) is identical to the covariance matrix of $\mathbf{F}$ as given by Eq. (58). The SNR for this task and observer is thus

$$[\mathrm{SNR(SKE/LBG, Hot, }\hat{\mathbf{F}})]^2 = \Delta\mathbf{F}^{\dagger}[\mathbf{K}_{\mathbf{F}} + \overline{\mathbf{J}}_N{}^{-1}]^{-1}\Delta\mathbf{F}\,. \tag{60}$$

If $\overline{\mathbf{J}}_N$ is diagonal, this figure of merit becomes

$$[\mathrm{SNR(SKE/LBG, Hot, }\hat{\mathbf{F}})]^2 = \sum_{k=1}^{N}\frac{|\Delta\overline{F}_k|^2\overline{J}_{kk}}{1 + W_k\overline{J}_{kk}}\cdot \tag{61}$$

The various figures of merit derived here are summarized in Table 1 and discussed in Sections 7 and 8 below.

6. ESTIMATION TASKS

As discussed in detail in paper I, a common estimation task is to estimate the integral of the object over a specified region of interest (ROI). In the present continuous formulation, the parameter to be estimated is the scalar θ defined by

$$\theta = \int_S \mathrm{d}\mathbf{r}\, t(\mathbf{r})f(\mathbf{r})\,, \tag{62}$$

where $t(\mathbf{r})$ is a template function that defines the ROI.

This parameter may not be estimable if the template contains null functions of the system operator $\mathcal{H}$, since many different true values of θ then lead to the same data and there is no way to construct an estimator that is unbiased for all values of the parameter. We can avoid

Table 1. Summary of Figures of Merit

Task	Observer	Figure of Merit	Form when $\mathbf{B}$ is Diagonal and $J_{kk} \simeq \beta_{kk}/\overline{g}_{\text{eff}}(k)$
Estimation of $\{F_k,\ k = 1 \ldots N\}$	Unbiased, efficient estimator	$\text{var}(\hat{F}_k) = (\mathbf{J}_n{}^{-1})_{kk} \geq \frac{1}{J_{kk}}$	$\text{var}(\hat{F}_k) = \frac{\overline{g}_{\text{eff}}(k)}{\beta_{kk}}$
SKE/BKE	Ideal/Hotelling access to $\mathbf{g}$	$\text{SNR}^2 = \Delta\mathbf{F}^\dagger \mathbf{J} \Delta\mathbf{F}$	$\sum_{k=1}^{\infty} \frac{\lvert\Delta F_k\rvert^2 \beta_{kk}}{\overline{g}_{\text{eff}}(k)}$
SKE/BKE	Ideal/Hotelling access to $\hat{\mathbf{F}}$ (efficient, unbiased)	$\Delta\mathbf{F}^\dagger \mathbf{J}_N \Delta\mathbf{F}$	$\sum_{k=1}^{N} \frac{\lvert\Delta F_k\rvert^2 \beta_{kk}}{\overline{g}_{\text{eff}}(k)}$
SKE/LBG	Hotelling access to $\hat{\mathbf{F}}$ (efficient, unbiased)	$\Delta\mathbf{F}^\dagger[\mathbf{K_F} + \overline{\mathbf{J}}_N{}^{-1}]^{-1}\Delta\mathbf{F}$	$\sum_{k=1}^{N} \frac{\lvert\Delta \overline{F}_k\rvert^2 \beta_{kk}}{\overline{g}_{\text{eff}}(k) + W_k \beta_{kk}}$
Estimation of activity in ROI	Linear ROI estimator access to $\hat{\mathbf{F}}$ (efficient, unbiased)	$\text{var}(\hat{\theta}) = \mathbf{T}^\dagger \mathbf{J}_N{}^{-1} \mathbf{T}$	$\sum_{k=1}^{N} \frac{\lvert T_k\rvert^2 \overline{g}_{\text{eff}}(k)}{\beta_{kk}}$

this difficulty by requiring that $t(\mathbf{r})$ be represented in the form

$$t(\mathbf{r}) = \sum_{k=1}^{N} T_k \Phi_k(\mathbf{r}) . \tag{63}$$

Since we have previously assumed that $\mathbf{B}_N$ is nonsingular, no Fourier coefficients for $k \leq N$ lie in the null space, and the θ constructed with this template is estimable. Note that the support of the template function is the same as the support of the object but that the template has fuzzy edges because of the finite upper limit in the sum over k.

Let us suppose that the observer wishing to estimate θ has available efficient, unbiased estimates of the Fourier coefficients $\hat{F}_k$, $k = 1 \ldots N$. If an efficient estimate exists, the maximum-likelihood estimate is efficient. Moreover, the maximum-likelihood estimate of a function $\theta(\mathbf{F})$ is $\theta(\hat{\mathbf{F}})$. Therefore we consider the estimator

$$\begin{aligned}\hat{\theta} &= \int_S \mathrm{d}\mathbf{r}\, t(\mathbf{r}) \sum_{k=1}^{N} \hat{F}_k \Phi_k(\mathbf{r}) \\ &= \int_S \mathrm{d}\mathbf{r} \sum_{k'=1}^{N} T_{k'}{}^* \Phi_{k'}{}^*(\mathbf{r}) \sum_{k=1}^{N} \hat{F}_k \Phi_k(\mathbf{r}) ,\end{aligned} \tag{64}$$

where the complex conjugates are legal since $t(\mathbf{r})$ is real and we have taken care to include in the sum components corresponding to both $\boldsymbol{\rho}_k$ and $-\boldsymbol{\rho}_k$.

At this point we can invoke orthogonality of the basis functions Φ_k, provided that the region of support is a cube and the Fourier grid satisfies the Nyquist condition. Under these conditions we have

$$\int_S \mathrm{d}\mathbf{r}\, \Phi_{k'}{}^*(\mathbf{r}) \Phi_k(\mathbf{r}) = V \delta_{kk'} , \tag{65}$$

where V is the volume of the region of support. For supports of other shapes, Eq. (65) is only an approximation. To obtain orthonormal basis functions, we must divide each Φ by $\sqrt{V}$, which is equivalent to incorporating a factor $1/V$ into Eq. (64). We then have

$$\hat{\theta} = \sum_{k=1}^{N} T_k{}^* \hat{F}_k = \mathbf{T}^\dagger \hat{\mathbf{F}} . \tag{66}$$

It is now straightforward to compute the mean and the variance of this estimator if we assume that $\hat{\mathbf{F}}$ is efficient and unbiased. We obtain

$$\langle \hat{\theta} \rangle = \left\langle \sum_{k=1}^{N} T_k{}^* \hat{F}_k \right\rangle = \mathbf{T}^\dagger \mathbf{F} = \theta , \tag{67}$$

$$\text{var}(\hat{\theta}) = \mathbf{T}^\dagger \mathbf{J}_N{}^{-1} \mathbf{T} , \tag{68}$$

where $\mathbf{F}$ is to be understood as an $N \times 1$ vector. If $\mathbf{J}_N$ is diagonal we find that

$$\text{var}(\hat{\theta}) = \sum_{k=1}^{N} \frac{|T_k|^2}{J_{kk}} . \tag{69}$$

This form is compared with other figures of merit in Table 1 and is discussed in more detail below.

7. LOCATION UNCERTAINTY AND SYSTEM OPTIMIZATION

All the tasks considered above lead to a figure of merit in the form $\mathbf{x}^\dagger \mathbf{A} \mathbf{x}$ (see the third column of Table 1). For estimation of an individual Fourier component, the vector $\mathbf{x}$ has a one in a single location and zeros elsewhere, and for ROI estimation, $\mathbf{x}$ is $\mathbf{T}$, the Fourier representation of the template defining the region. For the various classification tasks, $\mathbf{x}$ is the Fourier representation of the difference signal $\Delta\mathbf{F}$. Neither $\mathbf{T}$ nor $\Delta\mathbf{F}$ depends at all on the imaging system. The matrix $\mathbf{A}$, on the other hand, is related in all cases to the Fisher information matrix of the system (see Table 1). This factorization, like NEQ (see Section 8 below), thus separates out the specifics of the task from the system characterization, and it gives us a vehicle for understanding system optimization.

As noted in Section 4 [see relation (39)], the system is optimized for estimation of individual Fourier coefficients by making the Fisher information matrix diagonal and making the diagonal elements as large as possible. For Gaussian noise with constant variance or low-contrast Poisson noise, $\mathbf{J} \simeq \mathbf{B}/\sigma^2$, so the optimum system for the task of estimation of the Fourier coefficients is one designed according to the BG design rules, with a diagonal crosstalk matrix. For high-contrast Poisson noise, some improvement in the performance on this task could, in principle, be acheived by designing for diagonal $\mathbf{J}_N$ rather than for diagonal $\mathbf{B}_N$, but then the system would be optimized for one particular object since $\mathbf{J}_N$ depends on the object in that case.

One might expect that this same strategy, striving for diagonal $\mathbf{J}_N$ with large diagonal elements, would also yield an optimal system design for the other tasks considered. If we are trying to detect a signal or estimate a parameter on the basis of a data set consisting of estimates of Fourier coefficients, it is natural to expect that the strategy that optimizes the input to this procedure will optimize the performance of it. In other words, one might think that a design strategy that yields the best $\hat{\mathbf{F}}$ should also yield the best figure of merit for any task based on $\hat{\mathbf{F}}$.

This conclusion is partly correct. From an inspection of the last column of Table 1, we can immediately say that, if $\mathbf{J}_N$ is diagonal, it is advantageous to make each of the diagonal elements J_{kk} as large as possible. This will make the SNR's in the last column of the table large and the variances small, as we would like. The delicate point comes when $\mathbf{J}_N$ is not diagonal and we inquire whether the off-diagonal elements help or hurt the system performance for the task considered. In other words, what should one do to the system to optimize the more general figures of merit in the third column of the table?

A Hermitian form $\mathbf{x}^\dagger\mathbf{A}\mathbf{x}$ is not necessarily an extremum when $\mathbf{A}$ is diagonal. Addition of off-diagonal elements can either increase or decrease the value of the quantity, depending on the nature of $\mathbf{x}$ and the off-diagonal elements. One can easily construct examples, for all the tasks considered, in which a system with a nondiagonal $\mathbf{J}_N$ has better performance than one with a diagonal $\mathbf{J}_N$ having the same diagonal elements. For all the tasks considered so far, except for the elemental one of estimating the individual Fourier coefficients, off-diagonal elements can indeed be beneficial.

What does this mean physically? To answer this question, consider first the SKE/BKE detection task for a particular continuous difference signal $\Delta f_0(\mathbf{r})$, with Fourier representation $\Delta\mathbf{F}_0$. It will help to think of $\Delta f_0(\mathbf{r})$ as a spatially compact, even pointlike, object centered at point $\mathbf{r} = \mathbf{r}_0$. In a medical context this $\Delta f(\mathbf{r})$ might represent a small tumor *in a precisely known location*. An obvious strategy for building a system to detect this signal is to concentrate our detector resources on the known location. Mathematically, this means making all detector sensitivity functions $h_m(\mathbf{r})$ as large as possible at the fixed point $\mathbf{r}_0$. Suppose we could go so far as to make $h_m(\mathbf{r}) \simeq K\delta(\mathbf{r} - \mathbf{r}_0)$, where K is a constant. Then we would find [see Eqs. (6) and (7)] that

$$\Psi_{mk} = \int_S \mathrm{d}\mathbf{r} K\delta(\mathbf{r} - \mathbf{r}_0)\Phi_k(\mathbf{r}) = K \exp(2\pi i \boldsymbol{\rho}_k \cdot \mathbf{r}_0)\,. \quad (70)$$

The crosstalk matrix and the Fisher information matrix are then given by

$$\beta_{kk'} = MK^2 \exp[2\pi i(\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k'}) \cdot \mathbf{r}_0] \quad (71)$$

and

$$J_{kk'} = K^2 \exp[2\pi i(\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k'}) \cdot \mathbf{r}_0] \sum_{m=1}^{M} \frac{1}{{\sigma_m}^2}\,, \quad (72)$$

respectively. Both of these matrices are about as far from diagonal as they could be, since $|\beta_{kk'}|$ and $|J_{kk'}|$ are independent of both k and k', yet the system corresponding to these matrices is quite well suited for detection of a small signal at $\mathbf{r}_0$. The resulting SNR is high because the phase factors in $\mathbf{J}$ in Eq. (49) or $\mathbf{B}$ in Eq. (50) conspire to cancel those in $\Delta\mathbf{F}_0$ and ${\Delta\mathbf{F}_0}^\dagger$. The designer has cheated and optimized the system for one narrowly specified task without any concern for other tasks. Performance on the chosen task is optimized, but the resulting system cannot be used to reconstruct a good image, in a qualitative sense, or to obtain low-variance estimates of the Fourier coefficients. It will not perform well even on signal-detection tasks with the same signal in different locations.

The same kind of cheat could be used on any of the other tasks considered so far, since in each case the location of the signal or template is specified exactly.

To avoid this sort of overly specialized optimization, we must consider a range of tasks,[20] and one simple way to do so is to allow the signal or the ROI to be positioned anywhere in the region of support of the object. Consider the generic form for a figure of merit $\mathbf{x}^\dagger\mathbf{A}\mathbf{x}$ and let $\mathbf{x}_0$ be the Fourier representation for a signal (or template in the case of ROI estimation) centered at the origin. The SNR for a signal centered at $\mathbf{r}_0$ is given by

$$[\mathrm{SNR}(\mathbf{r}_0)]^2 = \sum_{k=1}^{N}\sum_{k'=1}^{N} {x_{0k}}^* A_{kk'} x_{0k'} \exp[2\pi i(\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k'}) \cdot \mathbf{r}_0]\,. \quad (73)$$

If we perform a large number of SKE detection tasks with the signal at different locations (all of which are known to the observer on each trial) and average the results, we find that

$$\begin{aligned}\langle \mathrm{SNR}^2\rangle &\equiv \frac{1}{V}\int_S \mathrm{d}\mathbf{r}_0[\mathrm{SNR}(\mathbf{r}_0)]^2 \\ &= \sum_{k=1}^{N}\sum_{k'=1}^{N} {x_{0k}}^* A_{kk'} x_{0k'} \frac{1}{V}\int_S \mathrm{d}\mathbf{r}_0 \exp[2\pi i(\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k'}) \cdot \mathbf{r}_0] \\ &\simeq \sum_{k=1}^{N} |x_{0k}|^2 A_{kk}\,, \end{aligned} \quad (74)$$

where we have used Eq. (65), the orthonormality of our Fourier basis functions. The approximation sign in this equation is a strict equality if the region of support is an appropriately chosen cube.

The key point from this discussion is that the variability in signal location makes the off-diagonal elements of $\mathbf{A}$ useless and leads to the same average performance (averaged over signal locations) as would be obtained from a diagonal matrix with the same diagonal elements. For signal detection in a nonrandom background, $\mathbf{A}$ is $\mathbf{J}_N$ or approximately $\mathbf{B}/\sigma^2$, so the strategy should be to make J_{kk} or β_{kk} large, irrespective of what happens to the off-diagonal elements.

For signal detection in a random background, however, $\mathbf{A}$ is $(\mathbf{K}_\mathbf{F} + {\overline{\mathbf{J}}_N}^{-1})^{-1}$, and it is not immediately obvious how one should choose $\overline{\mathbf{J}}_N$ to maximize the diagonal elements of this $\mathbf{A}$. The inequality derived in Appendix A, relation (A13), is the key to optimizing $\overline{\mathbf{J}}_N$ in this case.

Temporarily deleting the subscripts for clarity, we manipulate $(\mathbf{K} + \mathbf{J}^{-1})^{-1}$ into the form of inequality (A13) by transforming $\mathbf{K}$ into the identity matrix (whitening it) as follows:

$$(\mathbf{K} + \mathbf{J}^{-1})^{-1} = [\mathbf{K}^{1/2}(\mathbf{I} + \mathbf{K}^{-1/2}\mathbf{J}^{-1}\mathbf{K}^{-1/2})\mathbf{K}^{1/2}]^{-1} = \mathbf{K}^{-1/2}(\mathbf{I} + \mathbf{K}^{-1/2}\mathbf{J}^{-1}\mathbf{K}^{-1/2})^{-1}\mathbf{K}^{-1/2}. \quad (75)$$

For a general random background, the SNR after this averaging process is thus

$$\langle[\mathrm{SNR(SKE/RBG, Hot, }\hat{\mathbf{F}})]^2\rangle \le \sum_{k=1}^{N} \frac{|(\mathbf{K_F}^{-1/2}\Delta\mathbf{F})_k|^2}{1 + [\mathbf{K_F}^{-1/2}\overline{\mathbf{J}}_N{}^{-1}\mathbf{K_F}^{-1/2}]_{kk}}, \quad (76)$$

where we have used relations (57), (74), and (A13).

For the quasi-stationary LBG, $\mathbf{K_F}$ is diagonal in the Fourier representation, so we find that

$$\langle[\mathrm{SNR(SKE/LBG, Hot, }\hat{\mathbf{F}})]^2\rangle \le \sum_{k=1}^{N} \frac{|\Delta F_k|^2\overline{\mathbf{J}}_{kk}}{1 + [\mathbf{K_F}]_{kk}\overline{\mathbf{J}}_{kk}} \simeq \sum_{k=1}^{N} \frac{|\Delta F_k|^2\beta_{kk}}{\overline{g}_{\mathrm{eff}}(k) + [\mathbf{K_F}]_{kk}\beta_{kk}}, \quad (77)$$

where the last step holds for the moderate-contrast Poisson model. The SNR for extreme low-contrast Poisson noise or signal-independent Gaussian noise is obtained by replacement of $\overline{g}_{\mathrm{eff}}(k)$ with the constant σ^2.

Since inequality (A13) becomes an equality for a diagonal matrix, performance on this task is optimized for diagonal $\overline{\mathbf{J}}_N{}^{-1}$, which in turn requires diagonal $\overline{\mathbf{J}}_N$. In this case, because of the specific nature of the background random process, off-diagonal elements of $\mathbf{J}$ or $\mathbf{B}$ are detrimental, and the average SNR is maximized for a diagonal Fisher or crosstalk matrix.

8. CONNECTION WITH NOISE-EQUIVALENT QUANTA

To make contact with a significant portion of the literature on image quality, we now relate the results above to the concept of noise-equivalent quanta (NEQ). This concept was originally introduced by Shaw in a photographic context[39] and later extended to more general imaging systems.[40–42] Wagner and Brown[4] demonstrated the role played by NEQ in signal-detection theory and showed how it generalized for many of the 2D and 3D geometries used in medical imaging. Tapiovaara and Wagner[42] recently reviewed applications of the NEQ concept in medical imaging.

Consider a 2D LSIV imaging system and stationary Gaussian noise.[4] In that case the SNR for an ideal (or Hotelling) observer is given by

$$[\mathrm{SNR(SKE/BKE, LSIV, stat)}]^2 = \int_\infty \mathrm{d}^2\rho \left\{\frac{[G_0\ \mathrm{MTF}(\boldsymbol{\rho})]^2}{\mathrm{NPS}(\boldsymbol{\rho})}\right\}|\Delta F(\boldsymbol{\rho})|^2, \quad (78)$$

where $\boldsymbol{\rho}$ is the spatial-frequency vector (now 2D), G_0 is the low-frequency ($\boldsymbol{\rho} = 0$) input-to-output gain factor, $\mathrm{MTF}(\boldsymbol{\rho})$ is the modulation transfer function, $\Delta F(\boldsymbol{\rho})$ is the Fourier *transform* of the continuous-difference signal $f_2(\mathbf{r}) - f_1(\mathbf{r})$, and $\mathrm{NPS}(\boldsymbol{\rho})$ is the noise power spectrum. This result can be derived from Eqs. (47) and (48) by passing to the limit of continuous variables and realizing that the operator $\mathcal{H}$ is diagonal in the spatial-frequency domain for an LSIV system, and the covariance matrix is diagonal in this domain for stationary noise.

In Eq. (78) the integrand has been separated into two factors, one (in curly brackets) representing a combination of properties of the imaging system hardware and one, $|\Delta F(\boldsymbol{\rho})|^2$, representing the task. The quantity $G_0\ \mathrm{MTF}(\boldsymbol{\rho})$ may be thought of as a general input-to-output transfer factor whose square refers or scales the NPS back to the scale and units of the input, namely, those of $\Delta F(\boldsymbol{\rho})$. When the task is described on a relative scale (fractional change in activity, radiance, etc.), the appropriate combination of factors within the curly brackets has units of quanta (or counts) per unit area and is referred to as the number of noise-equivalent quanta, $\mathrm{NEQ}(\boldsymbol{\rho})$.

A related quantity is the generalized NEQ (GNEQ) introduced by Barrett *et al.*[43] in connection with the LBG problem. If the background is described by a 2D stationary Gaussian random process with power spectral density $W(\boldsymbol{\rho})$ and the imaging system is LSIV, the relevant SNR is

$$[\mathrm{SNR(SKE/LBG, LSIV, stat)}]^2 = \int_\infty \mathrm{d}^2\rho \left\{\frac{[G_0\ \mathrm{MTF}(\boldsymbol{\rho})]^2}{W(\boldsymbol{\rho})[G_0\ \mathrm{MTF}(\boldsymbol{\rho})]^2 + \mathrm{NPS}(\boldsymbol{\rho})}\right\}|\Delta F(\boldsymbol{\rho})|^2, \quad (79)$$

where $\mathrm{NPS}(\boldsymbol{\rho})$ is understood as being the power spectrum of the Poisson noise, conditional on a particular background but then averaged over all backgrounds so that the result is again stationary.[18,21,22] Under these conditions $\mathrm{NPS}(\boldsymbol{\rho})$ is a constant N_0 proportional to the mean background level, and the GNEQ is defined as

$$\mathrm{GNEQ}(\boldsymbol{\rho}) = \frac{[G_0\ \mathrm{MTF}(\boldsymbol{\rho})]^2}{W(\boldsymbol{\rho})[G_0\ \mathrm{MTF}(\boldsymbol{\rho})]^2 + N_0}. \quad (80)$$

An analogy to expressions in the last column of Table 1 is immediately apparent. If we interpret β_{kk} as the squared modulus of the absolute transfer function (including the gain factor G_0) and if the crosstalk matrix is diagonal, then the coefficient of $|\Delta F_k|^2$ in $[\mathrm{SNR(SKE/BKE, Hot, }\hat{\mathbf{F}})]^2$ or $[\mathrm{SNR(SKE/BKE, Hot, }\mathbf{g})]^2$ has the structure of $\mathrm{NEQ}(\boldsymbol{\rho})$. If we can approximate $\mathbf{J}$ as $\mathbf{B}/\sigma^2$, then the $\mathrm{NPS}(\boldsymbol{\rho})$ is the constant σ^2, analogous to white noise. If we use the moderate-contrast Poisson model, $\overline{g}_{\mathrm{eff}}(k)$ plays the role of $\mathrm{NPS}(\boldsymbol{\rho})$.

Similarly, the coefficient of $|\Delta F_k|^2$ in the last form of Eq. (77) is the analog of $\mathrm{QNEQ}(\boldsymbol{\rho})$ but again without any assumption of shift invariance of the system or of signal-independent noise.

All the analogies developed above are based on the last column of Table 1, which lists the special cases applicable when $\mathbf{B}$ is diagonal and the moderate-contrast Poisson model is used. Recall, however, the discussion in Section 7 where we noted that off-diagonal elements were irrelevant if we calculated the average over signal locations as in approximation (74). With that simple modification throughout, the expressions in the last column are applicable even if $\mathbf{B}$ is not diagonal. We need only consider the diagonal part of the matrices. In effect, the variability in signal location has forced a sort of stationarity on the expressions for the figures of merit. That means that NEQ and GNEQ are far more widely applicable than previously recognized. Expressions with exactly the same structure as the integrals in Eqs. (78) and (79) can be adduced for virtually any imaging system.

9. SUMMARY AND CONCLUSIONS

This paper has continued the program, initiated in paper I of the series, of calculating objective figures of merit for image quality. Both papers defined image quality on the basis of the performance of a mathematical observer on a specific task of practical interest. In both papers, both classification and estimation tasks were considered.

One difference from paper I is that in the current paper we adopted a fully continuous object model and treated the imaging system as a continuous-to-discrete mapping. Taking advantage of the finite support of real objects, we represented the continuous object as an infinite Fourier series multiplied by a support function. This representation was also used earlier by Barrett and Gifford[25] in a paper that we refer to as BG.

Another difference is that in this paper we focused on issues of system design, not reconstruction algorithms. The figures of merit were calculated either on the assumption that the observer had access to the raw, unreconstructed data or that the reconstruction consisted of efficient, unbiased estimates of the Fourier coefficients. In Appendix B we show that such estimates can be exactly realized by the Gauss–Markov estimator in the case of Gaussian noise and well approximated by the maximum-likelihood estimator with Poisson noise.

As in BG the system was described in this paper by a matrix called the crosstalk matrix. On purely geometric grounds BG argued that a reasonable goal for system design was to make this matrix as nearly diagonal as possible. In the present paper we have shown the relation between the BG design strategy and objective figures of merit. The unifying concept is the Fisher information matrix relevant to estimation of the Fourier coefficients. For both normal and Poisson noise models, we have shown that this Fisher matrix is closely related to the crosstalk matrix. If the measurements are independent and identically distributed, the Fisher and crosstalk matrices differ by a constant. If the measurements are statistically independent but not identically distributed, the Fisher matrix is similar to the crosstalk matrix except for an object-dependent weighting factor inside the sum [see Eq. (34) or (36)].

A full characterization of the imaging system requires the infinite crosstalk or the Fisher matrix, but only the $N \times N$ submatrix enters directly into the estimation of the first N Fourier coefficients. Nevertheless, as a practical matter, a system designer must ensure that no higher spatial frequencies are aliased into the band of interest, and for this purpose the full matrix should be investigated (step 3 in the BG design strategy). If this condition is satisfied, then one can use the finite Fisher information matrix to set Cramer–Rao lower bounds on the variances of estimates of the first N Fourier coefficients.

As shown by the examples in Subsection 2.E, it is often possible to control diagonal and off-diagonal elements of the crosstalk matrix more or less independently. If the Fisher and crosstalk matrices are proportional, as in the case of independent, identically distributed Gaussian noise, this statement obviously holds for the Fisher matrix as well. For Poisson noise the two matrices differ only by a weighting factor in the summand, so the examples in Subsection 2.E also show that diagonal and off-diagonal elements of the Fisher information matrix can often be controlled independently. In these circumstances the Cramer–Rao bounds on estimates of Fourier coefficients are at least approximately minimized by the BG design strategy. Of course, we cannot rule out the possibility that there may be practical design problems in which a reduction in off-diagonal elements is accompanied by a reduction in diagonal ones, and in those cases we need a task-specific scalar figure of merit.

Figures of merit were derived here for a variety of classification tasks. Included were detection of an exactly known signal superimposed on an exactly known background (SKE/BKE) and detection of an exactly known signal on a random background (SKE/RBG). The special case in which the RBG is described by a stationary random process truncated by the known object support was referred to as the lumpy-background (LBG) paradigm. The SKE/BKE problem has been investigated extensively in the literature and, especially for shift-invariant imaging and stationary noise, has led to the productive concept of NEQ. The LBG paradigm was introduced more recently to permit focus on the role of spatial resolution and other deterministic imaging properties. In this problem NEQ is naturally extended to what we call generalized NEQ (GNEQ).

In this paper we expressed the figures of merit for these tasks in terms of the Fisher and crosstalk matrices. The chosen observer for the classification tasks was the ideal linear discriminant, which we call the Hotelling observer, and we considered separately the cases in which this observer had access to the raw data or only to an image of finite spatial resolution. In the latter case we assumed that the image was formed from unbiased efficient estimates of the first N Fourier coefficients, so these coefficients were the data given to the observer.

With both data sets we found that the SNR for the SKE/BKE problem has the form $\Delta\mathbf{F}^{\dagger}\mathbf{J}\Delta\mathbf{F}$, which reduces to $\Delta\mathbf{F}^{\dagger}\mathbf{B}\Delta\mathbf{F}/\sigma^2$ for constant data variance, where $\Delta\mathbf{F}$ is the Fourier description of the signal to be detected [see Eqs. (49)–(53)]. For the observer with access only to the reconstructed image, $\mathbf{J}$ or $\mathbf{B}$ is the finite submatrix here. The observer with access to the raw data might perform better on this task simply because Fourier coefficients ΔF_k with $k > N$ might carry useful information.

The important point about this result, however, is that there is no obvious advantage to having a diagonal Fisher or crosstalk matrix. In other words, it is possible that whatever measures are needed to design the system for diagonal $\mathbf{J}$ or $\mathbf{B}$ would actually reduce performance on this task; the off-diagonal elements can contribute positively to SNR(SKE/BKE). Since off-diagonal elements are evidence of aliasing, this finding says that aliasing is not necessarily detrimental to an SKE/BKE task and may even be helpful.

A similar result came from consideration of the task of estimating the integral of the object over a specified region of interest (ROI). Again the figure of merit has the same algebraic structure, and again there is no obvious advantage to having a diagonal crosstalk matrix. Even for an estimation task the design strategy that leads to the best estimates of the individual Fourier coefficients does not produce the best estimate of a scalar parameter derived from the Fourier coefficients.

We resolved this paradox by considering the effects of variability in the location of the signal to be detected or in the ROI for the estimation task. We showed by a simple example how a system designer with foreknowledge of the precise location of the signal could concentrate all available detector resources on that area, optimizing the SKE figure of merit for that location at the expense of other desirable properties of the system, including the ability to estimate the Fourier coefficients.

To avoid this difficulty we suggested that, at least for purposes of system design, the SKE figure of merit should be averaged over all possible locations of the signal. This is equivalent to evaluating the system on the basis of a series of SKE detection experiments with the signal at different locations, all known to the observer in each experiment. This paradigm should not be confused with experiments in which the signal can be anywhere in the image and the observer's task is to determine, without knowledge of the location, whether the signal is present.

Averaging the SNR over signal locations has an important effect on the mathematical expressions for the figures of merit, all of which have the form $\mathbf{x}^\dagger\mathbf{A}\mathbf{x}$. The effect of location averaging is just to erase the off-diagonal terms of $\mathbf{A}$. Since $\mathbf{A}$ is usually the Fisher or crosstalk matrix (see Table 1), the resulting expressions are identical in form to ones obtained for the same tasks with a diagonal Fisher or crosstalk matrix. In other words, in most cases the off-diagonal elements are irrelevant with location variability.

An interesting exception to the last statement was found for the quasi-stationary LBG, where $\mathbf{A}$ contains a contribution from the background. Since this contribution, the intraclass scatter matrix for the object, was assumed to be diagonal in the Fourier representation, we found that the SNR is necessarily reduced by the presence of off-diagonal elements in $\mathbf{J}$ or $\mathbf{B}$. Since a stationary background is merely one in which there is no preferred location for features that might disrupt the detection task, it is not unreasonable to use this model for system design.

When the figures of merit were averaged over signal location, a striking parallel to the concept of NEQ emerged. Originally developed with the assumption of linear, shift-invariant imaging and stationary noise, NEQ has proven to be a useful tool for evaluating many kinds of imaging systems. Under these assumptions, the spatial-frequency domain is the natural one for description of both signal and noise, and in this domain the integrand in the SNR(SKE) expression factors neatly into a system-dependent part and a signal-dependent part. The system-dependent part, or NEQ, is the square of the system MTF divided by the noise power spectrum, and the signal-dependent part is just the squared modulus of the Fourier transform of the signal.

The NEQ concept was extended in previous literature to the case of SKE detection in a lumpy background. A similar factorization of the integrand resulted, with the system-dependent factor referred to as GNEQ, but the assumptions of shift invariance and stationarity were still required. Although more general matrix expressions that did not require these assumptions have been known for some time, they were not readily interpreted in terms of spatial frequencies and did not provide the same obvious factorization as NEQ and GNEQ.

After averaging over signal locations, all the SNR expressions (see the last column of Table 1) have precisely the same structure as SNR expressions based on NEQ and GNEQ. Both cases involve a sum (or integral) over the frequency domain, with the summand being a product of a system-dependent part and a signal-dependent part. For the SKE task the system-dependent part is the direct analog of $\mathrm{MTF}^2/\mathrm{NPS}$, namely, $\beta_{kk}/\overline{g}_{\mathrm{eff}}(k)$, and the signal-dependent part is again the squared modulus of the Fourier representation of the signal. This strong parallel to NEQ was obtained without any assumption of shift invariance or noise stationarity. A similar conclusion holds for GNEQ. Thus these useful concepts are much more widely applicable than had been previously recognized.

One way to understand this conclusion is to note that the Fourier domain is more useful in situations in which there is no preferred spatial origin. A shift-invariant system has no preferred origin, so the imaging operator is diagonal in the Fourier domain; Fourier transformation is equivalent to singular-value decomposition for such systems. Stationary noise has no preferred origin, so the noise covariance matrix is diagonal in the Fourier domain; Fourier transformation is a Karhunen–Loeve (KL) transformation for stationary noise. If both conditions prevail—shift-invariant imaging and stationary noise—then the Fourier domain is natural for both signal and noise and the NEQ factorization results. If one or both conditions fail, to achieve the factorization we would have to seek a transformation that is simultaneously a SVD and a KL. Although that procedure can, in principle, be done, and indeed was explored by Paxman and colleagues[44,45] for system optimization, it has the drawback that a different representation is needed for every system design considered.

Our approach of using a Fourier (series) representation in all cases avoids the latter difficulty, but it is in general neither a KL or a SVD. The trick of averaging over signal locations, however, removes the preferred spatial origin (except for the minor constraint of the object support) and makes the Fourier description much more widely useful. The resulting factorization should lead to new ways of looking at figures of merit for image quality.

APPENDIX A: PROOFS OF MATRIX INEQUALITIES

The second inequality in relation (29), which plays an important role in this paper, is a special case of an inequality given by Horn.[46] He proved it through use of Hadamard (element-by-element) products of matrices, but an independent proof that does not appeal to Hadamard products is given here.

Any positive-definite Hermitian $N \times N$ matrix $\mathbf{O}$ can be expanded in the form

$$\mathbf{O} = \sum_{n=1}^{N} \alpha_n \mathbf{u}_n \mathbf{u}_n{}^{\dagger}\,, \tag{A1}$$

where the eigenvalues $\{\alpha_n\}$ are real and positive and the eigenvectors $\{\mathbf{u}_n\}$ form a complete orthonormal set. If the eigenvalues are distinct, the orthonormality is easy to prove. If they are not, an orthonormal set can still be constructed by the Gram–Schmidt procedure, which we will presume to have been used.

The ith diagonal element of $\mathbf{O}$ is given by

$$O_{ii} = \sum_{n=1}^{N} \alpha_n |u_{ni}|^2\,, \tag{A2}$$

where u_{ni} denotes the ith component of the basis vector $\mathbf{u}_n$. The inverse of $\mathbf{O}$ is given by

$$\mathbf{O}^{-1} = \sum_{n=1}^{N} \frac{1}{\alpha_n} \mathbf{u}_n \mathbf{u}_n{}^{\dagger}\,, \tag{A3}$$

and its diagonal element is

$$[\mathbf{O}^{-1}]_{ii} = \sum_{n=1}^{N} \frac{1}{\alpha_n} |u_{ni}|^2\,. \tag{A4}$$

The Cauchy inequality states that, for real quantities $\{a_n\}$ and $\{b_n\}$,

$$\sum_{n=1}^{N} a_n{}^2 \sum_{n'=1}^{N} b_{n'}{}^2 \geq \left[\sum_{n=1}^{N} a_n b_n\right]^2. \tag{A5}$$

If $N > 1$, this inequality is an equality if and only if $a_n = \text{constant} \times b_n$.

To apply this inequality to the present problem, we choose $a_n = \sqrt{\alpha_n}\,|u_{ni}|$ and $b_n = |u_{ni}|/\sqrt{\alpha_n}$, with the result that

$$O_{ii}[\mathbf{O}^{-1}]_{ii} = \sum_{n=1}^{N} \alpha_n |u_{ni}|^2 \sum_{n'=1}^{N} \frac{1}{\alpha_{n'}} |u_{n'i}|^2 \geq \left[\sum_{n=1}^{N} |u_{ni}|^2\right]^2 = 1\,, \tag{A6}$$

where the last step follows from the completeness of the $\{\mathbf{u}_n\}$.

We next show that the equality in relation (A6) holds if and only if $\mathbf{O}$ is diagonal. The *if* is obvious since $u_{ni} = \delta_{ni}$ if $\mathbf{O}$ is diagonal, and there is then only one term in the sum over n, and the Cauchy inequality is trivially an equality. To show the *only if,* suppose that $\mathbf{O}$ is not diagonal. Then more than one term is required, and the Cauchy inequality becomes an equality if and only if $a_n = \sqrt{\alpha_n}\,|u_{ni}| = Kb_n = K|u_{ni}|/\sqrt{\alpha_n}$, which implies that $\alpha_n = K = \text{constant}$, which in turn implies that $\mathbf{O} = K\mathbf{I} =$ diagonal, in contradiction to our original assumption.

Thus we conclude that, for a positive-definite Hermitian matrix $\mathbf{O}$,

$$O_{ii}[\mathbf{O}^{-1}]_{ii} \geq 1\,, \tag{A7}$$

with equality if and only if $\mathbf{O}$ is diagonal.

The next inequality that we wish to consider involves diagonal elements of matrices of the form $(\mathbf{I} + \mathbf{O}^{-1})^{-1}$. This form occurs in the SKE problem in a stationary LBG [see Eqs. (60) and (75)]. As above, we assume that $\mathbf{O}$ is Hermitian and positive definite.

Equation (A1) and the orthogonality properties of the $\{\mathbf{u}_n\}$ lead to

$$(\mathbf{I} + \mathbf{O}^{-1})^{-1} = \sum_{n=1}^{N} \frac{\lambda_n}{1 + \lambda_n} \mathbf{u}_n \mathbf{u}_n{}^{\dagger}\,. \tag{A8}$$

The diagonal element of this matrix is given by

$$[(\mathbf{I} + \mathbf{O}^{-1})^{-1}]_{kk} = \sum_{n=1}^{N} \frac{\lambda_n}{1 + \lambda_n} |u_{nk}|^2\,. \tag{A9}$$

Now we make use of the fact that $f(t) \equiv t/(1 + t)$ is a concave function of t. Hence it satisfies

$$f[\alpha t_1 + (1 - \alpha)t_2] \geq \alpha f(t_1) + (1 - \alpha)f(t_2)\,, \qquad 0 \leq \alpha \leq 1\,. \tag{A10}$$

This result can be paraphrased by saying that a concave function defines a convex set of points. That is, if we define a set S of points in the $x-y$ plane by saying that $(x, y) \in S$ if $y \leq f(x)$, then inequality (A10) says that S is a convex set if $f(x)$ is a concave function.

An easy extension of inequality (A10) is Jensen's inequality,[47]

$$f\left(\sum_n \alpha_n t_n\right) \geq \sum_n \alpha_n f(t_n)\,, \qquad \sum_n \alpha_n = 1, \qquad \alpha_n \geq 1\,. \tag{A11}$$

We shall use this inequality with $t_n = \lambda_n$ and $\alpha_n = |u_{nk}|^2$, which satisfies the required normalization by the completeness of the $\{\mathbf{u}_n\}$. We then find that

$$\frac{\sum_n \lambda_n |u_{nk}|^2}{1 + \sum_n \lambda_n |u_{nk}|^2} \geq \sum_n \frac{\lambda_n}{1 + \lambda_n} |u_{nk}|^2\,. \tag{A12}$$

Since $\sum_n \lambda_n |u_{nk}|^2$ is just O_{kk}, we can also write that

$$[(\mathbf{I} + \mathbf{O}^{-1})^{-1}]_{kk} \leq \frac{O_{kk}}{1 + O_{kk}}\,, \tag{A13}$$

with equality if $\mathbf{O}$ is diagonal.

APPENDIX B: EFFICIENT ESTIMATORS

Since much of this paper is based on efficient, unbiased estimators of the Fourier coefficients, we look briefly at the form of such estimators in this appendix. We assume throughout that we have M measurements $\{g_m\}$ and wish to estimate N Fourier coefficients $\{F_n\}$, where $N \leq M$. The Fisher information matrix is thus the finite one $\mathbf{J}_N$, but we drop the subscript in this appendix. We assume that $\mathbf{J}$ is nonsingular, a condition that is easy to satisfy when $N \leq M$.

Critical to discussing efficient estimators is the score vector $\mathbf{s}(\mathbf{g}, \mathbf{f})$, which is the vector of derivatives of the log likelihood. The nth component of $\mathbf{s}(\mathbf{g}, \mathbf{f})$ is given by

$$s_n(\mathbf{g}, \mathbf{F}) = \frac{\partial}{\partial F_n} \ln p(\mathbf{g}|\mathbf{F}). \tag{B1}$$

The score is clearly a random vector since it depends on $\mathbf{g}$. Its mean is zero, and its covariance matrix is the Fisher information matrix $\mathbf{J}$.[26–28] An important theorem proved by Mardia *et al.*[27] states that a necessary and sufficient condition for an estimator to be efficient and unbiased is that it be an affine linear function of the score in the form

$$\hat{\mathbf{F}}_{\text{eff}} = \mathbf{F} + \mathbf{J}^{-1}\mathbf{s}(\mathbf{g}, \mathbf{F}), \tag{B2}$$

where the subscript eff stands for efficient. Any estimator in this form is automatically unbiased, since the mean of $\mathbf{s}(\mathbf{g}, \mathbf{F})$ is zero, and it is easy to show that it is efficient.

It is also known[26] that, if an efficient estimator exists, it must be the maximum-likelihood (ML) estimator. Furthermore, even if there is no exact efficient estimator, the ML estimator is asymptotically efficient and unbiased in the limit of a large number of independent data samples.[26,27]

Since there are no constraints such as positivity on the parameters $\mathbf{F}$, one obtains the ML estimator by setting all derivatives of the log likelihood to zero. In terms of the score, that means that

$$\mathbf{s}(\mathbf{g}, \hat{\mathbf{F}}_{\text{ML}}) = \mathbf{0}. \tag{B3}$$

These considerations are easy to apply in the case of multivariate Gaussian noise, where the score is given by

$$\mathbf{s}(\mathbf{g}, \mathbf{F}) = \mathbf{\Psi}^{\dagger}\mathbf{K}_{\epsilon}{}^{-1}(\mathbf{g} - \mathbf{\Psi F}). \tag{B4}$$

If we note that

$$\mathbf{J}^{-1} = (\mathbf{\Psi}^{\dagger}\mathbf{K}_{\epsilon}{}^{-1}\mathbf{\Psi})^{-1}, \tag{B5}$$

then the right-hand side of Eq. (B2) is independent of $\mathbf{F}$, and the efficient, unbiased estimator of $\mathbf{F}$ is given by

$$\hat{\mathbf{F}}_{\text{eff}} = (\mathbf{\Psi}^{\dagger}\mathbf{K}_{\epsilon}{}^{-1}\mathbf{\Psi})^{-1}\mathbf{\Psi}^{\dagger}\mathbf{K}_{\epsilon}{}^{-1}\mathbf{g}. \tag{B6}$$

This is essentially the classical Gauss–Markov estimator, discussed, for example, in paper I,[18] but with the usual matrix transpose replaced by an adjoint since here we are discussing complex parameters. It is easy to verify that $\hat{\mathbf{F}}_{\text{eff}} = \hat{\mathbf{F}}_{\text{ML}}$ in this case. Thus, for multivariate Gaussian noise, an efficient, unbiased estimator exists and is given by the Gauss–Markov estimator or, equivalently, the ML estimator. For the special case where $N = M$ and $\mathbf{\Psi}^{-1}$ exists, $\hat{\mathbf{F}}_{\text{eff}} = \hat{\mathbf{F}}_{\text{ML}} = \mathbf{\Psi}^{-1}\mathbf{g}$.

The problem is a bit more subtle for Poisson noise. One obtains the nth component of the score in the Poisson case by differentiating the logarithm of Eq. (35) with respect to F_n, with the result that

$$s_n(\mathbf{g}, \mathbf{F}) = \sum_{m=1}^{M} \Psi_{mn}{}^{*}\left[\frac{g_m}{(\mathbf{\Psi F})_m} - 1\right]. \tag{B7}$$

To put this result in a convenient vector form, we define the ratio of two vectors on a Hadamard or component-by-component basis, so that $\mathbf{c} = \mathbf{a}/\mathbf{b}$ is interpreted to mean that $c_m = a_m/b_m$. With this convention Eq. (B7) becomes

$$\mathbf{s}(\mathbf{g}, \mathbf{F}) = \mathbf{\Psi}^{\dagger}\left[\frac{\mathbf{g}}{\mathbf{\Psi F}} - \mathbf{1}\right], \tag{B8}$$

where $\mathbf{1}$ is an $M \times 1$ vector of all ones.

The ML estimate must satisfy Eq. (B3), which in this case becomes

$$\mathbf{s}(\mathbf{g}, \hat{\mathbf{F}}_{\text{ML}}) = \mathbf{\Psi}^{\dagger}\left[\frac{\mathbf{g}}{\mathbf{\Psi}\hat{\mathbf{F}}_{\text{ML}}} - 1\right] = \mathbf{0}. \tag{B9}$$

If $N = M$ and $\mathbf{\Psi}^{-1}$ exists, then this equation is satisfied by $\hat{\mathbf{F}}_{\text{ML}} = \mathbf{\Psi}^{-1}\mathbf{g}$, and it is easy to show that this estimate is also efficient. If $N < M$, however, $\mathbf{\Psi}^{-1}$ does not exist, and we cannot in general find any estimate for which $\mathbf{\Psi}\hat{\mathbf{F}} = \mathbf{g}$. On the other hand, the operator $\mathbf{\Psi}^{\dagger}$ has a null space, so all we require is that $\mathbf{g}/\mathbf{\Psi}\hat{\mathbf{F}}_{\text{ML}} - \mathbf{1}$ lie in that null space. Different realizations of the data vector $\mathbf{g}$ lead to different null vectors. In spite of the null space of $\mathbf{\Psi}^{\dagger}$, Eq. (B9) uniquely specifies $\hat{\mathbf{F}}$ if the finite crosstalk matrix $\mathbf{\Psi}^{\dagger}\mathbf{\Psi}$ (or $\mathbf{B}_N$) is invertible, so that $\mathbf{\Psi}$ has full column rank.

Various iterative search algorithms, including a variant of the expectation-maximization algorithm, can be devised for solving Eq. (B9). One approach to computing the covariance matrix of $\hat{\mathbf{F}}_{\text{ML}}$ would be to choose one of these algorithms and develop a recursion relation for the random component of the estimate at each iteration. This approach, which has been applied to the expectation-maximization algorithm by Barrett *et al.*,[48] yields expressions for the mean and covariance of $\hat{\mathbf{F}}_{\text{ML}}$ (and even its full multivariate probability-density function) as a function of iteration number. A Monte Carlo evaluation of this theory performed by Wilson *et al.*[49] shows that it gives excellent predictions of all these statistical properties when the expectation-maximization algorithm is applied to ML estimation of voxel values. To our knowledge, ML methods have not previously been applied to the estimation of Fourier coefficients.

ML estimation of Fourier coefficients is actually significantly easier to analyze than estimation of voxel values since we do not have to account for the positivity constraint. This allows us to derive an approximate expression for the covariance matrix of $\hat{\mathbf{F}}_{\text{ML}}$ without considering a specific algorithm. As in Ref. 48, we expand $\hat{\mathbf{F}}_{\text{ML}}$ in powers of deviations from its mean and retain only first-order terms. In that reference the mean of $\hat{\mathbf{F}}$ varied with iteration number, thus requiring a recursive theory, but here we take advantage of the consistency of the ML estimator, which requires that $\hat{\mathbf{F}}_{\text{ML}}$ approach $\mathbf{F}$ as the total number of measured photons approaches infinity. In this limit, $\mathbf{g} = \mathbf{\Psi F}$, so Eq. (B9) is satisfied by $\hat{\mathbf{F}}_{\text{ML}} = \mathbf{F}$. Denoting $\hat{\mathbf{F}}_{\text{ML}} - \mathbf{F}$ by $\delta\hat{\mathbf{F}}$ and recalling that $\mathbf{g} = \mathbf{\Psi F} + \boldsymbol{\epsilon}$, we have

$$\frac{\mathbf{g}}{\mathbf{\Psi}\hat{\mathbf{F}}_{\text{ML}}} - \mathbf{1} = \frac{\boldsymbol{\epsilon}}{\mathbf{\Psi F}} - \frac{\mathbf{\Psi}\delta\hat{\mathbf{F}}_{\text{ML}}}{\mathbf{\Psi F}} + \ldots, \tag{B10}$$

where the additional terms are powers or products of $\delta\hat{\mathbf{F}}$ and $\boldsymbol{\epsilon}$. We note that

$$\mathbf{\Psi}^{\dagger}\left[\frac{\mathbf{\Psi}\delta\hat{\mathbf{F}}}{\mathbf{\Psi F}}\right] = \mathbf{\Psi}^{\dagger}\ \text{diag}\left(\frac{\mathbf{1}}{\mathbf{\Psi F}}\right)\mathbf{\Psi}\delta\hat{\mathbf{F}} = \mathbf{J}\delta\hat{\mathbf{F}}, \tag{B11}$$

where $\text{diag}(\mathbf{1}/\mathbf{\Psi F})$ is a diagonal matrix with element mm given by $1/(\mathbf{\Psi F})_m$. Combining Eq. (B10) truncated to linear terms with Eq. (B11), we find that

$$\delta\hat{\mathbf{F}} = \mathbf{J}^{-1}\mathbf{\Psi}^{\dagger}\left[\frac{\boldsymbol{\epsilon}}{\mathbf{\Psi F}}\right]. \tag{B12}$$

Since we can easily show that $\mathbf{\Psi}^{\dagger}\mathbf{1} = \mathbf{JF}$, Eq. (B12) is equivalent to Eq. (B2). Alternatively, a direct calculation of the covariance matrix from Eq. (B12) yields $\mathbf{J}^{-1}$.

The conclusion is that $\hat{\mathbf{F}}_{\mathrm{ML}}$ is efficient as long as we can make the approximation of dropping higher-order terms in Eq. (B10), an approximation that improves as more photons are accumulated.

We could have stated this result simply as a consequence of the theorem that an ML estimate is asymptotically efficient, but the derivation given above has two advantages. First, it shows how to interpret the term asymptotically. Usual statements of the theorem assume that a large number of independent data vectors are used, while the treatment here requires only that $\boldsymbol{\epsilon}/\mathbf{\Psi F}$ be small, which it is as $\mathbf{g}$ becomes large. In a sense, for the case of Poisson noise, every photon is an independent sample.

The second advantage is that we can estimate the magnitude of the error in the approximation by examining the size of the neglected terms. A detailed treatment of these errors is beyond the scope of this paper, but we note that the same basic approximation proved remarkably robust, even with relatively few photons, in the Monte Carlo investigations of Ref. 49.

ACKNOWLEDGMENTS

The authors have benefited greatly from conversations with Craig Abbey, Howard Gifford, Don Wilson, Edward Soares, and Tim White. The manuscript has been critically read by Gabor Herman, Al Hero, Jannick Rolland, Marie Kijewski, Steven Moore, Markku Tapiovaara, and Eric Hansen; their comments have been invaluable in extending and clarifying many points. This research was supported in part by the National Institutes of Health under grants PO1 CA23417 and RO1 CA52643.

Correspondence should be addressed to Harrison H. Barrett, Department of Radiology, University of Arizona, Tucson, Arizona 85724. Tel: 602-626-6815, e-mail: barrett@zen.radiology.arizona.edu.

REFERENCES

1. G. E. Backus and J. F. Gilbert, "Numerical application of a formalism for geophysical inverse problems," Geophys. J. R. Astron. Soc. **13,** 247–276 (1967).
2. M. Bertero, C. DeMol, and E. R. Pike, "Linear inverse problems with discrete data. I: General formulation and singular system analysis," Inv. Prob. **1,** 301–330 (1985).
3. R. F. Wagner, K. J. Myers, D. G. Brown, M. J. Tapiovaara, and A. E. Burgess, "Higher-order tasks: human vs. machine performance," in *Medical Imaging III: Image Formation,* S. J. Dwyer, R. G. Jost, and R. H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090,** 183–194 (1989).
4. R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," Phys. Med. Biol. **30,** 489–518 (1985).
5. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual discrimination," Science **214,** 93–94 (1981).
6. J. A. Swets, "Measuring the accuracy of diagnostic systems," Science **240,** 1285–1293 (1988); J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems* (Academic, New York, 1982).
7. K. M. Hanson, "Variations in task and the ideal observer," in *Application of Optical Instrumentation in Medicine XI (Atlanta),* G. D. Fullerton, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **419,** 60–67 (1983).
8. K. M. Hanson, "Optimization for object localization of the constrained algebraic reconstruction technique," in *Medical Imaging III: Image Formation,* S. J. Dwyer, R. G. Jost, and R. H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090,** 146–153 (1989).
9. S. P. Mueller, M. F. Kijewski, S. C. Moore, and B. L. Holman, "Maximum-likelihood estimation—a model for optimal quantitation in nuclear medicine," J. Nucl. Med. **31,** 1693–1701 (1990).
10. M. F. Kijewski, "The Barankin bound: a model of detection with location uncertainty," in *Mathematical Methods in Medical Imaging,* D. C. Wilson and J. N. Wilson, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1768,** 153–160 (1992).
11. B. M. W. Tsui, C. E. Metz, F. B. Atkins, S. J. Starr, and R. N. Beck, "A comparison of optimum spatial resolution in nuclear imaging based on statistical theory and on observer performance," Phys. Med. Biol. **23,** 654–676 (1978).
12. C. F. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21,** 720–733 (1986).
13. W. E. Smith and H. H. Barrett, "Hotelling trace criterion as a figure of merit for the optimization of imaging systems," J. Opt. Soc. Am. A **3,** 717–725 (1986).
14. G. T. Herman and K. T. D. Yeung, "Evaluators of image reconstruction algorithms," Int. J. Imag. Syst. Technol. **1,** 187–195 (1989).
15. G. T. Herman and D. Odhner, "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography," IEEE Trans. Med. Imag. **10,** 336–346 (1991).
16. A. E. Hero and L. Shao, "Information analysis of single photon computed tomography with count losses," IEEE Trans. Med. Imag. **9,** 117–127 (1990).
17. H. H. Barrett, T. A. Gooley, K. A. Girodias, J. P. Rolland, T. A. White, and J. Yao, "Linear discriminants and image quality," Image Vision Comput. **10,** 451–460 (1992).
18. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," J. Opt. Soc. Am. A **7,** 1266–1278 (1990).
19. R. F. Wagner, D. G. Brown, and C. E. Metz, "On the multiplex advantage of coded source/aperture photon imaging," in *Digital Radiography,* W. R. Brody, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **314,** 72–76 (1981).
20. 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 **7,** 1279–1293 (1990).
21. J. P. Rolland, "Factors influencing lesion detection in medical imaging," Ph.D. dissertation (University of Arizona, Tucson, Arizona, 1990).
22. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9,** 649–658 (1992).
23. N E Hartsough, H. H. Barrett, H. B. Barber, and J. M. Woolfenden, "Predicting the performance of gamma-ray imaging devices in the task of tumor detection," in *Proceedings of the 1991 IEEE Medical Imaging Conference, Orlando, Fla.* (Institute of Electrical and Electronics Engineers, New York, 1992), pp. 2113–2117).
24. K. J. Myers, R. F. Wagner, K. M. Hanson, H. H. Barrett, and J. P. Rolland, "Human and quasi-Bayesian observers of images limited by quantum noise, object variability, and artifacts," in *Medical Imaging 1994: Image Perception,* H. L. Kundel, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **2166,** 180–190 (1994).
25. H. H. Barrett and H. C. Gifford, "Cone–beam tomography with discrete data sets," Phys. Med. Biol. **39,** 451–476 (1994).
26. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), Vol. 1.
27. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis* (Academic, San Diego, Calif., 1979).
28. M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook* (American Sciences, Columbus, Ohio, 1985).
29. H. H. Barrett, "The Radon transform and its applications," in

Progress in Optics, E. Wolf, ed. (North-Holland, Amsterdam, 1984), Vol. 21, pp. 219–285.

30. J. Pilz, *Bayesian Estimation and Experimental Design in Linear Regression Models* (Wiley, Bergakademie Freiberg, Germany, 1991).
31. A. O. Hero and J. A. Fessler, "A fast recursive algorithm for computing CR-type bounds for image reconstruction problems," presented at the Institute of Electrical and Electronics Engineers Nuclear Science Symposium, Orlando, Fla., October 1992.
32. R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "The Hotelling trace criterion and its correlation with human observer performance," J. Opt. Soc. Am. A **4,** 945–953 (1987).
33. H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," Proc. Natl. Acad. Sci. (USA) **90,** 9758–9765 (1993).
34. K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2,** 1752–1759 (1985).
35. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4,** 2447–2457 (1987).
36. J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," in *Mathematical Methods in Medical Imaging,* D. C. Wilson and J. N. Wilson, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1768,** 161–168 (1992).
37. J. Yao, "Predicting human performance by model observers," Ph.D. dissertation (University of Arizona, Tucson, Ariz., 1994).
38. A. B. Whalen, *Detection of Signals in Noise* (Academic, New York, 1971).
39. R. Shaw, "The equivalent quantum efficiency of the photographic process," J. Photog. Sci. **11,** 199–204 (1963).
40. R. Shaw, "Evaluating the efficiency of imaging processes," Rep. Prog. Phys. (UK) **41,** 1103–1155 (1978).
41. J. C. Dainty and R. Shaw, *Image Science* (Academic, London, 1974).
42. M. J. Tapiovaara and R. F. Wagner, "SNR and noise measurements for medical imaging: I. A practical approach based on statistical decision theory," Phys. Med. Biol. **38,** 71–92 (1993).
43. H. H. Barrett, J. P. Rolland, R. F. Wagner, and K. J. Myers, "Detection of known signals in inhomogeneous, random backgrounds," in *Medical Imaging III: Image Formation,* S. J. Dwyer, R. G. Jost, and R. H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090,** 176–182 (1989).
44. R. G. Paxman, "Coordinated design of restoration algorithm and coded aperture," Ph.D. dissertation (University of Arizona, Tucson, Ariz., 1984).
45. R. G. Paxman, H. H. Barrett, W. E. Smith, and T. D. Milster, "Image reconstruction from coded data: code design," J. Opt. Soc. Am. A **2,** 501–509 (1985).
46. R. A. Horn, "The Hadamard product," Proc. Symp. Appl. Math. **40,** 87–169 (1990).
47. L. Rade and B. Westergren, *Beta Mathematics Handbook* (CRC, Boca Raton, Fla., 1989).
48. H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm: I. Theory," Phys. Med. Biol. **39,** 833–846 (1994).
49. D. W. Wilson, B. M. W. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm: II. Monte Carlo simulations," Phys. Med. Biol. **39,** 847–872 (1994).