# Objective assessment of image quality: effects of quantum noise and object variability

**Harrison H. Barrett**

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

Received October 10, 1989; accepted February 23, 1990

A number of task-specific approaches to the assessment of image quality are treated. Both estimation and classification tasks are considered, but only linear estimators or classifiers are permitted. Performance on these tasks is limited by both quantum noise and object variability, and the effects of postprocessing or image-reconstruction algorithms are explicitly included. The results are expressed as signal-to-noise ratios (SNR's). The interrelationships among these SNR's are considered, and an SNR for a classification task is expressed as the SNR for a related estimation task times four factors. These factors show the effects of signal size and contrast, conspicuity of the signal, bias in the estimation task, and noise correlation. Ways of choosing and calculating appropriate SNR's for system evaluation and optimization are also discussed.

## 1. INTRODUCTION

Image processing and image reconstruction are, naturally, concerned with making better images, but the word "better" is clearly defined only rarely in papers on these subjects. Instead, most papers simply present a new algorithm and one or two examples of images produced by it, relying entirely on subjective impression for assessment of the quality of the images.

There are, of course, exceptions to this indictment. For example, some quadratic measure such as the mean-square difference between the object and the image is frequently used as a figure of merit. This approach has serious difficulties in practice since all subjects in the real world are continuous, while all digital images consist of a finite set of numbers. There is thus an infinite number of objects equally consistent with any particular image; the choice of one of them for the calculation of mean-square error is highly arbitrary.

Two approaches to removing this arbitrariness are used. The first is to compare the processed image not with an actual object but with some discrete version of one, such as the digital representation used to simulate the data set. This approach tests the self-consistency of the algorithm but does not address its performance with real objects. Indeed, there are many algorithms that work well with simulated data from discrete objects but fail completely with real data from continuous objects.

A somewhat better approach is to consider an ensemble of objects and to calculate the average mean-square error, where the averaging process includes both the noise in the data and the variability in the objects themselves. This viewpoint leads to Wiener–Helstrom filtering, or some generalization of it, but it definitely does not lead to unambiguously optimal images.

The basic difficulty with both subjective assessment and mean-square error as measures of image quality is that they do not take into account the purpose of the image. A scientific or medical image is always produced for some specific purpose or task, and the only meaningful measure of its quality is how well it fulfills that purpose. An objective approach to assessment of image quality must therefore start with a specification of the task and then determine quantitatively how well the task is performed.

There are two generic kinds of task to be considered: classification and estimation. Classification is the usual task in medical imaging, where the ultimate purpose is to make a diagnosis. The physician making the diagnosis views the image (or set of images) and classifies it either as normal or as indicating some particular disease state. Some measure of the accuracy of the diagnosis is then the measure of image quality.

A simple example of classification is detection of, say, a tumor. Here the classes are tumor-present or tumor-absent, and the quality measure must be related to probability of detection and probability of false alarm. The trade-off between these two probabilities is quantified in the receiver operating characteristic (ROC) curve, originally devised for radar signal detection but applied with vigor to medical imaging.[1–4] Useful figures of merit include the area under the ROC curve and various detectability indices such as $d'$ and $d_a$.[3]

On the other hand, it is becoming increasingly common in medicine to acquire an image for the purpose of determining the value of some quantitative parameter of interest. A cardiologist might, for example, want to know how much blood is pumped out of the heart on each stroke, or a neurologist might want to know how much of a particular radioactive tracer is taken up in a certain region of the brain. Here the image and its associated processing algorithm may be treated as any other measuring instrument, with accuracy and precision of the measurement as indicators of quality. Recent examples of the use of an estimation task for assessment of image quality in radiology include research by Hanson[5,6] and Mueller *et al.*[7]

Similar considerations apply to other scientific imaging situations as well. In astronomical speckle interferometry, for example, one might want to detect the presence of a small

0740-3232/90/071266-13$02.00 

companion star near a brighter star, or one might want to estimate the relative magnitudes of the two stars. The first problem is a detection or classification task, while the second is an estimation task.

Performance of either estimation or classification tasks may be limited by many factors, but in this paper we shall concentrate on just two fundamental ones: quantum noise and object variability. Quantum noise is the ultimate limitation in any measurement system, and this limitation can be closely approached in practice in many applications in radiology and astronomy. On the other hand, there is always some intrinsic randomness in the objects being imaged. One manifestation of object variability is background clutter in a scene, which can limit the detectability of a signal of interest. Thus an imaging system is not necessarily quantum limited even if there is no other source of measurement noise, and object variability must be taken into account in a correct calculation of performance.

The goal of this paper is to discuss various possible figures of merit for both classification and estimation tasks and to show how they are influenced by quantum noise and by variability in the original objects. In addition, we shall derive some useful relationships between classification and estimation metrics.

One difficulty in this endeavor is that image quality, as we define it, depends not only on the image and the task but also on how we go about performing the task. For a classification task, different quality metrics would be derived for classification by a human observer or by some sort of machine classifier or pattern-recognition system. The best performance, by virtually any measure, would be obtained by the ideal Bayesian classifier. Similarly, an estimation task can be performed by *ad hoc* methods, by a maximum-likelihood estimator, or by a Bayesian estimator. The difficulty with Bayesian methods is that we almost never know the required probability densities.

To make the problem tractable, we shall restrict attention to *linear* operations. Classification will be assumed to be performed by computing a test statistic that is a linear combination of the image pixel values and comparing it with a threshold, and all estimators will likewise be linear. Except in certain cases, which we shall note as they occur, this means that the estimator or classifier is not optimal. Nevertheless, linear methods are often used in practice, and it is of considerable interest to assess performance with them. Even with so complex a classifier as the human being, linear models have had considerable success in predicting the performance. Though this paper is not primarily about performance of human observers, some comments about the relation between humans and our linear models will be made in Section 6.

In Section 2 we set up the general framework that we need and discuss the statistical models of the objects and the data. For generality, we suppose that the initial data set is not immediately usable and that some kind of processing or reconstruction step is needed before we have a usable image; some of the results will be simplified if this step is also linear.

In Section 3 we consider various approaches to estimation and derive the corresponding figures of merit. Three estimators will be considered: a simple integration of gray values in a region of interest, the optimal linear or Gauss–Markov estimator,[8] and the generalized Wiener estimator, which assumes some knowledge of object statistics. All three are linear.

A similar plan for classification problems is executed in Section 4. Three linear classification strategies are considered. The first forms a test statistic by simple matched filtering or integrating over a template; this test statistic is the classification counterpart of the region-of-interest estimator. Then we treat a prewhitening matched filter, the counterpart of the Gauss–Markov estimator. Finally, we discuss the optimum linear classifier due to Hotelling,[9] which is the counterpart of the Wiener estimator.

In Section 5 we discuss the interrelationship between classification and estimation and present some potentially useful factorization formulas. Some of the practical problems in applying these methods of assessment and in choosing a figure of merit are discussed in Section 6.

## 2. ASSUMPTIONS AND MODELS

A general linear imaging system can be described by

$$\mathbf{g} = H\mathbf{f} + \mathbf{n}, \tag{1}$$

where $\mathbf{f}$ represents the object, $\mathbf{g}$ is the data set, $H$ is the system operator, and $\mathbf{n}$ is the noise in the data. There is no loss of generality in writing the noise as additive so long as we recognize that the statistical properties of the noise may depend on $\mathbf{f}$. In a discrete model of the object, $\mathbf{f}$ is an $L \times 1$ column vector, $\mathbf{g}$ and $\mathbf{n}$ are $M \times 1$ vectors, and $H$ is an $M \times L$ matrix. It is, however, also permissible to regard $\mathbf{f}$ as a continuous object and $H$ as a linear operator mapping from the continuous object space to the discrete data space. We shall use the discrete vector description for $\mathbf{f}$ throughout, but, in fact, all results are applicable to the case of continuous objects simply by letting $L$ approach infinity. We shall always presume that the data set is digital, so $\mathbf{g}$ is always a discrete vector with a finite number of elements. No matter the size of $H$, we shall not assume that its inverse exists.

Note that there are two random vectors, $\mathbf{f}$ and $\mathbf{n}$, that contribute to the randomness in $\mathbf{g}$. [The special case of a nonrandom $\mathbf{f}$ can be handled by letting its probability-density function (PDF) approach an $L$-dimensional delta function.] We must therefore define several kinds of statistical average. Since the statistical properties of $\mathbf{n}$ depend on $\mathbf{f}$, in general, we define a conditional average over all realizations of $\mathbf{n}$ for a fixed $\mathbf{f}$ as

$$\langle \ldots \rangle_{n|f} \equiv \int \mathrm{d}\mathbf{n} p(\mathbf{n}|\mathbf{f}) \ldots, \tag{2}$$

where the ellipsis denotes the quantity to be averaged, $p(\mathbf{n}|\mathbf{f})$ is the conditional PDF of $\mathbf{n}$, given $\mathbf{f}$, and the integral is over all values for all components of $\mathbf{n}$. An average over $\mathbf{f}$ itself is given by

$$\langle \ldots \rangle_f \equiv \int \mathrm{d}\mathbf{f} p(\mathbf{f}) \ldots, \tag{3}$$

where $p(\mathbf{f})$ is the PDF of $\mathbf{f}$.

In classification problems, we are also interested in averages over only those objects in a given class. The average over the $j$th class is defined by

$$\langle \ldots \rangle_j \equiv \int d\mathbf{f} p(\mathbf{f}|j) \ldots, \tag{4}$$

where $p(\mathbf{f}|j)$ is the PDF for **f**, given that it belongs to the $j$th class. The overall average over **f** is given in terms of the class averages by

$$\langle \ldots \rangle_f = \sum_{j=1}^{J} P_j \langle \ldots \rangle_j, \tag{5}$$

where $P_j$ is the prior probability of occurrence of class $j$, with $\Sigma_{j=1}^{J} P_j = 1$.

### Noise Statistics

If there were no object variability, and the only noise were due to the discrete nature of the radiation, it would almost always be correct to take each component of **g** as an independent Poisson random variable. Exceptions to this statement could arise, for example, in optical imaging of a thermal source if the integration time were short compared with the coherence time, but this condition is hard to obtain in practice. The Poisson model is certainly valid for medical imaging with x rays or gamma rays.

It must be emphasized, however, that **g** is only conditionally Poisson, i.e., $P(g_i|\mathbf{f})$ is Poisson with mean $[H\mathbf{f}]_i$, with the subscript denoting the $i$th component of the vector. Thus the conditional mean of **g** is

$$\langle \mathbf{g} \rangle_{n|f} = H\mathbf{f}, \tag{6}$$

where an average of a vector quantity stands for the set of averages of the components. Since the components of **g** are conditionally independent, the conditional covariance matrix of the noise vector **n** is given by

$$K_{n|f} \equiv \langle \mathbf{n}\mathbf{n}^t \rangle_{n|f} = \mathrm{diag}(H\mathbf{f}), \tag{7}$$

where $\mathbf{n}\mathbf{n}^t$ is the outer or tensor product of **n** with itself (i.e., $[\mathbf{n}\mathbf{n}^t]_{ij} = n_i n_j$) and diag($H\mathbf{f}$) denotes a diagonal matrix, where the $i$th element along the diagonal is the $i$th component of the vector $H\mathbf{f}$.

The overall covariance matrix of **n**, denoted by $K_n$, is obtained by averaging the conditional covariance matrix over **f**:

$$K_n = \langle K_{n|f} \rangle_f = \mathrm{diag}(H\bar{\mathbf{f}}), \tag{8}$$

where $\bar{\mathbf{f}} = \langle \mathbf{f} \rangle_f$ is the overall mean of **f**.

Even though $K_n$ is diagonal, it does not follow that fluctuations in the components of **g** are uncorrelated. The covariance matrix for **g** itself is defined by

$$K_g \equiv \langle [\mathbf{g} - H\bar{\mathbf{f}}][\mathbf{g} - H\bar{\mathbf{f}}]^t \rangle_{n,f}$$

$$= \langle \langle [H(\mathbf{f} - \bar{\mathbf{f}}) + \mathbf{n}][H(\mathbf{f} - \bar{\mathbf{f}}) + \mathbf{n}]^t \rangle_{n|f} \rangle_f = HK_fH^t + K_n. \tag{9}$$

This form shows explicitly that the overall covariance matrix is the sum of two terms, one representing the object variability and the other the quantum noise. Even though the quantum-noise term $K_n$ is diagonal, the object-variability term $HK_fH^t$ is not. Physically, if a particular object pixel is brighter than its average, because of a chance fluctuation in **f**, then all components of **g** that are coupled to it by nonzero elements of $H$ will fluctuate upward together. (We assume here that all elements of $H$ are nonnegative.) Conversely, if that object pixel is dimmer than its average, then the same elements of **g** will fluctuate downward together. In either case, there is a positive correlation. On the other hand, if one observes these elements of **g** during repeated measurements on one realization of **f**, then, by the independent nature of Poisson random variables, the correlations are zero. In summary, elements of **g** are conditionally uncorrelated (for a given **f**), but the fluctuations in **f** produce correlations in **g**.

### Object Class Statistics

In the discussion of classification problems, we shall need to know the first- and second-order statistics for **f**, given that it belongs to a certain class. The class mean for the $j$th class is

$$\bar{\mathbf{f}}_j \equiv \langle \mathbf{f} \rangle_j, \tag{10}$$

while the class covariance matrix is

$$K_j \equiv \langle (\mathbf{f} - \bar{\mathbf{f}}_j)(\mathbf{f} - \bar{\mathbf{f}}_j)^t \rangle_j. \tag{11}$$

The separability of the classes is often discussed in terms of two scatter matrices, $S_1$ and $S_2$. The intraclass scatter matrix $S_2$ is simply the average covariance matrix, defined as

$$S_2 \equiv \sum_{j=1}^{J} P_j K_j, \tag{12}$$

where $P_j$ is again the prior probability of class $j$. The interclass scatter matrix $S_1$, which measures how far the class means for the data values deviate from their grand mean $\bar{\mathbf{f}}$, is defined as

$$S_1 \equiv \sum_{j=1}^{J} P_j (\bar{\mathbf{f}} - \bar{\mathbf{f}}_j)(\bar{\mathbf{f}} - \bar{\mathbf{f}}_j)^t, \tag{13}$$

where

$$\bar{\mathbf{f}} = \sum_{j=1}^{J} P_j \bar{\mathbf{f}}_j. \tag{14}$$

If there are $L$ object pixels, both $S_1$ and $S_2$ are $L \times L$ matrices. Since $S_2$ represents an ensemble covariance matrix, it will usually also have rank $L$. The rank of $S_1$, on the other hand, is much less than $L$; in fact it is simply $J - 1$, where $J$ is the number of classes.[10] Thus, for a two-class problem, $S_1$ has rank one, and it can be written as a single outer product:

$$S_1 = P_1 P_2 (\bar{\mathbf{f}}_2 - \bar{\mathbf{f}}_1)(\bar{\mathbf{f}}_2 - \bar{\mathbf{f}}_1)^t. \tag{15}$$

The overall covariance matrix for **f**, which is defined as

$$K_f \equiv \langle (\mathbf{f} - \bar{\mathbf{f}})(\mathbf{f} - \bar{\mathbf{f}})^t \rangle_f, \tag{16}$$

is given in terms of the scatter matrices[10] by

$$K_f = S_1 + S_2. \tag{17}$$

### Reconstruction Statistics

As indicated in Section 1, we assume for generality that we do not work directly with the raw data **g** but rather with some reconstructed image or estimate of **f**; we refer to this estimate as $\hat{\mathbf{f}}$ and regard it as an $N \times 1$ column vector, where $N$ is not necessarily the same as $L$, which is the number of components in **f**.

Of course, in an estimation task we might want to estimate some other quantity rather than $\mathbf{f}$ itself, but usually $\hat{\mathbf{f}}$ will be obtained as an intermediate step. The operator that forms $\hat{\mathbf{f}}$ from $\mathbf{g}$ will be denoted by $O$, so that we can write

$$\hat{\mathbf{f}} = O\{\mathbf{g}\} = O\{H\mathbf{f} + \mathbf{n}\} = OH\mathbf{f} + O\mathbf{n}, \tag{18}$$

where the last step holds only if $O$ is linear, in which case $O$ is an $N \times M$ matrix. In this case, we can also write

$$\hat{\mathbf{f}} = \mathbf{f} + [OH - I]\mathbf{f} + O\mathbf{n} = \mathbf{f} + \mathbf{b} + \mathbf{m}, \tag{19}$$

where $I$ is the unit operator, $\mathbf{b}$ is the bias, and $\mathbf{m}$ ($=O\mathbf{n}$) is the noise in the reconstruction. The term $\mathbf{m}$ differs from $\mathbf{b}$ in that it is the part that has zero mean when averaged over different noise realizations for a given $\mathbf{f}$:

$$\langle \mathbf{m} \rangle_{n|f} = 0. \tag{20}$$

The covariance matrix for $\mathbf{m}$ is the $N \times N$ matrix defined by

$$K_m \equiv \langle \mathbf{m}\mathbf{m}^t \rangle_{n,f} = OK_nO^t. \tag{21}$$

The bias $\mathbf{b}$, conversely, is that part of $\hat{\mathbf{f}} - \mathbf{f}$ that does not average conditionally to zero:

$$\mathbf{b} \equiv [OH - I]\mathbf{f} = \langle \hat{\mathbf{f}} - \mathbf{f} \rangle_{n|f}. \tag{22}$$

One important kind of bias in a reconstruction is due to the limited spatial resolution of the system, but other kinds of artifact or system imperfection may also produce bias. In a tomographic problem, for example, there is an inherent bias due to the limited number of angular samples.

Whatever its source, the bias will be described by a matrix $B$, also $N \times N$, which we call the bias matrix. It is defined by

$$B \equiv \langle \mathbf{b}\mathbf{b}^t \rangle_f. \tag{23}$$

The overall mean of the reconstruction is

$$\bar{\hat{\mathbf{f}}} \equiv \langle \hat{\mathbf{f}} \rangle_{n,f}, \tag{24}$$

while the overall covariance matrix is

$$\hat{K} \equiv \langle (\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}})(\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}})^t \rangle_{n,f} = OHK_fH^tO^t + K_m. \tag{25}$$

The notation here may be somewhat confusing; $\hat{K}$ is not an estimate of the covariance matrix but rather a covariance matrix of the estimate $\hat{\mathbf{f}}$. The main virtue of this notation is that it avoids ornaments on subscripts.

### Ensemble Mean-Square Error

As was mentioned in Section 1, the Wiener–Helstrom approach to image restoration[11] is to minimize the ensemble mean-square error (EMSE) in the reconstruction. For later comparison we give the expression for the EMSE in our notation:

$$\begin{aligned} \mathrm{EMSE}[\hat{\mathbf{f}}] \equiv \langle |\hat{\mathbf{f}} - \mathbf{f}|^2 \rangle_{n,f} &= \mathrm{tr}\langle (\hat{\mathbf{f}} - \mathbf{f})(\hat{\mathbf{f}} - \mathbf{f})^t \rangle_{n,f} \\ &= \mathrm{tr}\, B + \mathrm{tr}\, K_m, \end{aligned} \tag{26}$$

where the vertical bars denote the $L_2$ norm, tr denotes the trace of a matrix, and we have used the identity $\mathbf{u}^t\mathbf{v} = \mathrm{tr}(\mathbf{u}\mathbf{v}^t)$. There are thus two contributions to the EMSE, the bias and the stochastic noise $\mathbf{m}$.

### Reconstruction Class Statistics

The class mean of $\hat{\mathbf{f}}$ for the $j$th class is

$$\bar{\hat{\mathbf{f}}}_j \equiv \langle \hat{\mathbf{f}} \rangle_{n,j}, \tag{27}$$

and the corresponding class covariance matrix is

$$\hat{K}_j \equiv \langle (\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}}_j)(\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}}_j)^t \rangle_{n|j}. \tag{28}$$

Similarly, the scatter matrices for $\hat{\mathbf{f}}$ are

$$\hat{S}_2 = \sum_{j=1}^{J} P_j\hat{K}_j, \tag{29}$$

$$\hat{S}_1 = \sum_{j=1}^{J} P_j(\bar{\hat{\mathbf{f}}} - \bar{\hat{\mathbf{f}}}_j)(\bar{\hat{\mathbf{f}}} - \bar{\hat{\mathbf{f}}}_j)^t, \tag{30}$$

where again the ˆ indicates that we are dealing with scatter matrices for $\hat{\mathbf{f}}$. In terms of these scatter matrices, the overall covariance matrix for $\hat{\mathbf{f}}$ is

$$\hat{K} = \hat{S}_1 + \hat{S}_2. \tag{31}$$

If $O$ is linear, the scatter matrices for $\hat{\mathbf{f}}$ are related to those for $\mathbf{f}$ by

$$\hat{S}_1 = OHS_1H^tO^t, \tag{32}$$

$$\hat{S}_2 = OHS_2H^tO^t + OK_nO^t. \tag{33}$$

This form for $\hat{S}_2$ is again a sum of two terms, the first representing the object variability and the second representing the quantum noise. Note that $\hat{S}_1$, since it depends on only the class means, does not have the quantum-noise term.

## 3. ESTIMATION METHODS AND FIGURES OF MERIT

Let us suppose that we wish to estimate some scalar parameter of the form

$$\theta = \mathbf{w}^t\mathbf{f}, \tag{34}$$

where both $\mathbf{f}$ and $\mathbf{w}$ are $L \times 1$ column vectors and hence $\mathbf{w}^t$ is a $1 \times L$ row vector. For example, $\mathbf{w}$ might have a certain cluster of elements that are equal to one and the rest to zero. It then defines a region of interest (ROI), and $\theta$ represents the total strength of the object $\mathbf{f}$ over this ROI. Since $L$ may be very large, even approaching infinity, $\theta$ can represent a continuous integral with arbitrary weighting over some region in the object.

It is not obvious *a priori* that it is even meaningful to talk about estimating $\theta$. Suppose, for example, that the ROI is very small, even smaller than the size of a pixel in the reconstruction $\hat{\mathbf{f}}$. Then there could be a great deal of fine structure in $\mathbf{f}$ that was totally absent from $\hat{\mathbf{f}}$, and valid information about $\theta$ could not be obtained from $\hat{\mathbf{f}}$. To state the problem more technically, the imaging operator $H$ can have null functions, and any two objects $\mathbf{f}$ that differ by a null function must produce the same data set $\mathbf{g}$ and hence the same $\hat{\mathbf{f}}$ and the same value for any estimator of $\theta$. We cannot, in principle, evaluate how successful the estimator is since we have no way of knowing which of the infinite set of possible objects was actually present.

There are two ways to deal with this dilemma. The first is to restrict $\theta$ to be "estimable,"[8] which means that it is possible to find an unbiased linear estimator of $\mathbf{w}^t\mathbf{f}$ for all $\mathbf{f}$. This condition requires that

$$H^+H\mathbf{w} = \mathbf{w}, \tag{35}$$

where $H^+$ denotes the Moore–Penrose pseudoinverse of $H$

(which is defined even if $H$ is a continuous-to-discrete mapping) and $H^+H$ is therefore the operator that eliminates null functions. Thus $\theta$ is estimable if the template $\mathbf{w}$ has no null functions. More practically, this means that we can estimate integrals of $\mathbf{f}$ over regions that are large compared with the system resolution but not over smaller regions.

If $\theta$ is estimable from $\mathbf{g}$, it follows that one can find a vector $\mathbf{y}$ such that

$$\hat{\theta} = \mathbf{y}^t\mathbf{g} \tag{36}$$

and

$$\langle\hat{\theta}\rangle_{n|f} = \theta \qquad \text{for all } \mathbf{f}. \tag{37}$$

Similarly, if we want $\theta$ to be estimable directly from $\hat{\mathbf{f}}$ rather than from $\mathbf{g}$, we must require that

$$A^+A\mathbf{w} = \mathbf{w}, \tag{38}$$

where $A$ is the overall (linear) operator connecting $\mathbf{f}$ to $\hat{\mathbf{f}}$, i.e.,

$$A = OH. \tag{39}$$

Of course, the other way to deal with the problem of null functions and the resulting difficulty in defining a meaningful, unbiased estimator is to ignore it! In that case, one obtains a bias in $\hat{\theta}$ that can have any value, depending on what null functions are contained in $\mathbf{f}$. Nevertheless, one can define an average bias, which involves an ensemble average over $\mathbf{f}$; if this average bias is not too large, a biased estimator can be tolerated.

**Simple Region-of-Interest Estimation**

A suboptimal but common approach to estimation of $\theta$ is simply to apply the template $\mathbf{w}$ directly to the reconstructed image:

$$\hat{\theta}_{\mathrm{ROI}} = \mathbf{w}^t\hat{\mathbf{f}}. \tag{40}$$

This estimator, which we shall refer to as the ROI estimator, is indeed biased, with the bias given by

$$b_{\mathrm{ROI}} \equiv \langle\hat{\theta}_{\mathrm{ROI}}\rangle_{n|f} - \theta = \mathbf{w}^t\mathbf{b}, \tag{41}$$

where $\mathbf{b}$ is the bias in $\hat{\mathbf{f}}$ as defined in Eq. (22). The mean-square bias is given by

$$\langle[\mathbf{w}^t\mathbf{b}]^2\rangle_f = \mathrm{tr}[BW], \tag{42}$$

where $B$ is defined by Eq. (23) and

$$W \equiv \mathbf{w}\mathbf{w}^t. \tag{43}$$

The variance of this estimator, derived in Appendix A, is given by

$$\mathrm{var}(\hat{\theta}_{\mathrm{ROI}}) \equiv \langle(\hat{\theta}_{\mathrm{ROI}} - \bar{\hat{\theta}}_{\mathrm{ROI}})^2\rangle_{n,f} = \mathrm{tr}[\hat{K}W]. \tag{44}$$

This variance does not take into account the bias, so a better figure of merit is the EMSE of the estimate (also derived in Appendix A), given by

$$\mathrm{EMSE}[\hat{\theta}_{\mathrm{ROI}}] = \langle(\hat{\theta}_{\mathrm{ROI}} - \theta)^2\rangle_{n,f} = \mathrm{tr}(BW) + \mathrm{tr}(K_mW). \tag{45}$$

Note that the earlier expression for EMSE[$\hat{\mathbf{f}}$] in Eq. (26) is a special case of this one with $W$ replaced by the unit matrix. The present expression, however, has the advantage that the matrix $W$ concentrates the error calculation on the region that is important for a particular task; errors outside this region do not influence EMSE[$\hat{\theta}_{\mathrm{ROI}}$].

It is convenient to define a figure of merit as the reciprocal of an EMSE and to normalize it so that the absolute object strength is irrelevant. We thus define a signal-to-noise ratio (SNR) for any estimator as

$$[\mathrm{SNR}_{\mathrm{est}}]^2 \equiv \frac{\langle\theta^2\rangle_{n,f}}{\mathrm{EMSE}[\hat{\theta}]}. \tag{46}$$

For the simple ROI estimator, we have

$$[\mathrm{SNR}_{\mathrm{ROI}}]^2 = \frac{\mathrm{tr}[FW]}{\mathrm{tr}[BW] + \mathrm{tr}[K_mW]}, \tag{47}$$

where a mean-square measure of object strength is given by the matrix

$$F \equiv \langle\mathbf{f}\mathbf{f}^t\rangle_f. \tag{48}$$

**Pixel Signal-to-Noise Ratio**

A special case of $\mathrm{SNR}_{\mathrm{ROI}}$ is the commonly used pixel or point SNR, which is the relative standard deviation of $\hat{\mathbf{f}}$ at one point. Usually the theory of the pixel SNR is developed without consideration of either bias in $\hat{\mathbf{f}}$ or object variability. Thus the usual definition of $\mathrm{SNR}_{\mathrm{pixel}}$ for the $k$th pixel is

$$[\mathrm{SNR}_{\mathrm{pixel}}]^2 = \frac{[f_k]^2}{[K_{m|f}]_{kk}}, \tag{49}$$

where $[K_{m|f}]_{kk}$ (the $kk$ element of the matrix) is simply the conditional variance of $\hat{\mathbf{f}}$ at pixel $k$. This SNR thus depends on the object $\mathbf{f}$ as well as on the pixel being considered. Methods of calculating this quantity are discussed in detail by Barrett and Swindell.[12]

For comparison with the present paper, we generalize the definition of pixel SNR to include object variability and bias. Then the expression for $\mathrm{SNR}_{\mathrm{pixel}}$ is exactly the same as for $\mathrm{SNR}_{\mathrm{ROI}}$ in Eq. (47) but with the ROI shrunk to one pixel, so that

$$W_{ij} = \delta_{ik}\delta_{jk}. \tag{50}$$

With a larger ROI, we can relate $\mathrm{SNR}_{\mathrm{pixel}}$ to $\mathrm{SNR}_{\mathrm{ROI}}$ simply by preprocessing $\hat{\mathbf{f}}$ by convolution with the template $\mathbf{w}$; then $\mathrm{SNR}_{\mathrm{pixel}}$, at some point in the convolved image, is identical to $\mathrm{SNR}_{\mathrm{ROI}}$ *for a region centered at the same point in the original image*. The methods described by Barrett and Swindell for calculating $\mathrm{SNR}_{\mathrm{pixel}}$ are thus equally applicable to the calculation of $\mathrm{SNR}_{\mathrm{ROI}}$.

**Gauss–Markov Estimator**

Next we consider unbiased estimators and, in particular, the one that is best in the sense of giving the smallest conditional variance. This estimator, known as the best linear unbiased estimator (BLUE) or Gauss–Markov estimator, is also the maximum-likelihood estimator if we can assume that the noise is normally distributed.

If we have access to $\hat{\mathbf{f}} = OH\mathbf{f} + \mathbf{m}$ but not to the raw data $\mathbf{g}$, one useful form for the Gauss–Markov estimator of $\theta$ is[8,13]

$$\hat{\theta}_{\mathrm{GM}} = \mathbf{w}^t[A^tK_m{}^{-1}A]^+A^tK_m{}^{-1}\hat{\mathbf{f}}, \tag{51}$$

where again $A = OH$ and the plus denotes pseudoinverse.

It is always possible to factor a square, positive-definite matrix such as $K_m^{-1}$ into the product of two identical matrices, thus formally defining $K_m^{-1/2}$. Another useful form of the Gauss–Markov estimator can then be obtained by letting $C = K_m^{-1/2}A$ and using the identity (Ref. 8, p. 26)

$$C^+ = [C^tC]^+C^t. \tag{52}$$

This leads to

$$\hat{\theta}_{\mathrm{GM}} = \mathbf{w}^t[K_m^{-1/2}A]^+K_m^{-1/2}\hat{\mathbf{f}} \equiv \mathbf{v}^t\hat{\mathbf{f}}. \tag{53}$$

An interpretation of this form is that one should first remove the noise correlations in $\hat{\mathbf{f}}$, which is accomplished by the *prewhitening* filter $K_m^{-1/2}$, then perform a pseudoinverse operation on the result, and finally integrate over the template $\mathbf{w}$.

The Gauss–Markov estimator minimizes the conditional variance of $\hat{\theta}$ for a given $\mathbf{f}$. In this paper, however, we also allow $\mathbf{f}$ to be a random variable, so that the EMSE is the appropriate metric. In Appendix B we show that the EMSE is given by

$$\mathrm{EMSE}[\hat{\theta}_{\mathrm{GM}}] = \mathrm{tr}[VK_m] = \mathrm{tr}[(A^tK_m^{-1}A)^+W], \tag{54}$$

where $V \equiv \mathbf{v}\mathbf{v}^t$. Therefore

$$[\mathrm{SNR}_{\mathrm{GM}}]^2 = \frac{\mathrm{tr}[FW]}{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}. \tag{55}$$

### Wiener Estimator

Among all linear estimators, the Wiener (or Wiener–Helstrom) estimator is the one that minimizes the EMSE. It is also optimal among all estimators, in this same sense, if the object and the noise obey Gaussian statistics.

To apply the Wiener estimator, we must know the mean and covariance matrix of the object $\mathbf{f}$. For our problem, in which we have access only to $\hat{\mathbf{f}}$, the Wiener estimator has the well-known[13,14] form:

$$\hat{\theta}_{\mathrm{WE}} = \mathbf{w}^t(A^tK_m^{-1}A + K_f^{-1})^{-1}A^tK_m^{-1}(\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}}) + \mathbf{w}^t\bar{\mathbf{f}}. \tag{56}$$

If $K_f$ becomes large, so that we have no useful knowledge of the object statistics, the term $K_f^{-1}$ becomes negligible compared to $A^tK_m^{-1}A$ so long as we do not consider null components of the latter. Thus the inverse in Eq. (56) limits to the pseudoinverse, and the Wiener estimator approaches the Gauss–Markov estimator as in Eq. (51).

The EMSE for the Wiener estimator is given by

$$\mathrm{EMSE}[\hat{\theta}_{\mathrm{WE}}] = \mathrm{tr}[(A^tK_m^{-1}A + K_f^{-1})^{-1}W], \tag{57}$$

and the corresponding $\mathrm{SNR}^2$ is

$$[\mathrm{SNR}_{\mathrm{WE}}]^2 = \frac{\mathrm{tr}[FW]}{\mathrm{tr}[(A^tK_m^{-1}A + K_f^{-1})^{-1}W]}. \tag{58}$$

## 4. CLASSIFICATION METHODS AND FIGURES OF MERIT

For simplicity, we restrict attention here to binary (two-class) classification problems, but this does not entail any loss of generality since a multiclass problem can be formulated as a sequence of binary problems.

In a binary problem, some scalar test statistic $\lambda$ is formed from the data and compared with a threshold. If $\lambda$ exceeds the threshold the decision is, say, class 2, but otherwise it is class 1. Here we consider only linear test statistics of the general form

$$\lambda = \mathbf{u}^t\hat{\mathbf{f}}. \tag{59}$$

A common figure of merit[13] for binary classification problems is the detectability index $d$ (also known as $d_a$), defined by

$$d^2 = \frac{[\langle\lambda\rangle_{j=2} - \langle\lambda\rangle_{j=1}]^2}{P_1\,\mathrm{var}(\lambda|j=1) + P_2\,\mathrm{var}(\lambda|j=2)}, \tag{60}$$

where $\mathrm{var}(\lambda|j = i)$ denotes the conditional variance of $\lambda$ under hypothesis $i$. The index $d$ can be related to the area under the ROC curve, so it fulfills our requirement for a metric that relates to the probability of correct classification. For later convenience, however, we incorporate a scale factor $P_1P_2$ in the definition of our figure of merit, so that we define

$$[\mathrm{SNR}_\lambda]^2 = P_1P_2d^2. \tag{61}$$

Thus, for the common case where $P_1 = P_2 = 1/2$, $\mathrm{SNR}_\lambda$ is $d/2$.

It is straightforward to evaluate this $\mathrm{SNR}_\lambda$ in terms of the scatter matrices previously defined. The numerator in Eq. (60) is

$$\begin{aligned}[\langle\lambda\rangle_{j=2} - \langle\lambda\rangle_{j=1}]^2 &= [\mathbf{u}^t(\bar{\hat{\mathbf{f}}}_2 - \bar{\hat{\mathbf{f}}}_1)]^2 = \mathbf{u}^t(\bar{\hat{\mathbf{f}}}_2 - \bar{\hat{\mathbf{f}}}_1)(\bar{\hat{\mathbf{f}}}_2 - \bar{\hat{\mathbf{f}}}_1)^t\mathbf{u} \\ &= \mathrm{tr}[U(\bar{\hat{\mathbf{f}}}_2 - \bar{\hat{\mathbf{f}}}_1)(\bar{\hat{\mathbf{f}}}_2 - \bar{\hat{\mathbf{f}}}_1)^t] = [P_1P_2]^{-1}\,\mathrm{tr}[U\hat{S}_1],\end{aligned} \tag{62}$$

where

$$U \equiv \mathbf{u}\mathbf{u}^t. \tag{63}$$

The variances in the denominator of $\mathrm{SNR}_\lambda{}^2$ are given by

$$\mathrm{var}(\lambda|j) = \mathbf{u}^t\hat{K}_j\mathbf{u}. \tag{64}$$

The average covariance matrix is simply $\hat{S}_2$, so, finally,

$$[\mathrm{SNR}_\lambda]^2 = \frac{\mathrm{tr}[U\hat{S}_1]}{\mathrm{tr}[U\hat{S}_2]}. \tag{65}$$

This expression will be evaluated for several particular test statistics in the following sections.

### Simple Matched Filter

It is well known that detection of a nonrandom signal $\mathbf{s}$ in white, Gaussian noise is optimally performed with a simple matched filter of the form[13,15]

$$\lambda_{\mathrm{MF}} = \mathbf{s}^t\hat{\mathbf{f}}. \tag{66}$$

Under the same assumptions, this form is also optimal for discrimination between two nonrandom signals, but $\mathbf{s}$ is then to be interpreted as the difference between the signals.

It has been conjectured,[16] however, that the human observer uses a test statistic of this form even in more complicated circumstances where the noise is nonwhite or non-Gaussian or where the signal itself is random. Thus we evaluate here the performance of the matched filter in circumstances where it is decidedly suboptimal. From Eq. (65) we have immediately

$$[\mathrm{SNR}_{\mathrm{MF}}]^2 = \frac{\mathrm{tr}[S\hat{S}_1]}{\mathrm{tr}[S\hat{S}_2]}, \tag{67}$$

where

$$S \equiv \mathbf{s}\mathbf{s}^t. \tag{68}$$

To get back to a more familiar form, we assume that the filter $\mathbf{s}$ is actually $\bar{\bar{\hat{\mathbf{f}}}}_2 - \bar{\bar{\hat{\mathbf{f}}}}_1$, so that $S$ is identical (within an irrelevant constant) to $\hat{S}_1$. We also note that $\hat{S}_1$ or $S$ has rank one, so that the trace of its square is the square of its trace and, therefore,

$$[\mathrm{SNR}_{\mathrm{MF}}]^2 = \frac{[\mathrm{tr}(\hat{S}_1)]^2}{\mathrm{tr}[\hat{S}_1\hat{S}_2]}. \tag{69}$$

If we further assume that there is no object variability so that $\hat{S}_2$ is the same as $K_m$, we get

$$[\mathrm{SNR}_{\mathrm{MF}}]^2 = \frac{[\mathrm{tr}(\hat{S}_1)]^2}{\mathrm{tr}[\hat{S}_1K_m]}. \tag{70}$$

If the noise is stationary, $K_m$ is a Toeplitz (or block Toeplitz) matrix, often well approximated by a circulant matrix and hence diagonalizable by a discrete Fourier transform.[11] This means that the traces in Eq. (70) are easily performed in the Fourier domain, and this equation is in fact identical to the expression for the performance of a nonprewhitening matched filter as given by Wagner *et al.*[17] We emphasize, however, that several assumptions were necessary to get from the general result [Eq. (67)] to the Wagner form. Specifically, we had to assume that the signal was nonrandom and known exactly, that there was no variability in $\hat{\mathbf{f}}$ except that due to the random noise $\mathbf{m}$, and that the noise was stationary.

**Prewhitening Matched Filter**

If the noise is Gaussian but correlated, the optimum strategy for detection of a nonrandom signal $\mathbf{s}$ is to use the prewhitening matched filter[15] given by

$$\lambda_{\mathrm{GM}} = \mathbf{s}^t K_m^{-1}\hat{\mathbf{f}}. \tag{71}$$

An interesting way to rewrite this expression is

$$\lambda_{\mathrm{GM}} = [K_m^{-1/2}\mathbf{s}]^t K_m^{-1/2}\hat{\mathbf{f}}. \tag{72}$$

Note the similarity of this form to that of the Gauss–Markov estimator in Eq. (53). Again we first prewhiten with the operator $K_m^{-1/2}$, but now we follow that step with a matched filter for the prewhitened signal $K_m^{-1/2}\mathbf{s}$, whereas in the estimation case we had to do a pseudoinverse.

Once again we evaluate the performance of this filter in situations for which it is not optimal, i.e., when the object is variable with or without the signal. One rationale for doing this is that the observer might have some knowledge of the noise covariance matrix $K_m$, for example from a theoretical calculation, but might not know the full statistics of $\hat{\mathbf{f}}$. Thus the prewhitening matched filter would be expected to be an improvement over the nonprewhitening one, even though it is still not the optimal linear test statistic. From Eq. (65) we find that

$$[\mathrm{SNR}_{\mathrm{PW}}]^2 = \frac{\mathrm{tr}[K_m^{-1}\hat{S}_1K_m^{-1}S]}{\mathrm{tr}[K_m^{-1}\hat{S}_2K_m^{-1}S]}. \tag{73}$$

If the template $\mathbf{s}$ is actually the mean difference signal, so that (within a constant) $S = \hat{S}_1$, we get

$$[\mathrm{SNR}_{\mathrm{PW}}]^2 = \frac{\{\mathrm{tr}[K_m^{-1}\hat{S}_1]\}^2}{\mathrm{tr}[K_m^{-1}\hat{S}_2K_m^{-1}\hat{S}_1]}. \tag{74}$$

If we assume that there is no object variability so that $\hat{S}_2 = K_m$, then the conditions for which the filter is optimal (signal nonrandom and known exactly, noise described by $K_m$ alone) actually prevail, and we find that

$$[\mathrm{SNR}_{\mathrm{PW}}]^2 = \mathrm{tr}(K_m^{-1}\hat{S}_1). \tag{75}$$

With the further assumption of stationary noise, this expression is equivalent to the expression given by Wagner *et al.*[17] for the performance of a prewhitening matched filter.

**Optimum Linear Discriminant**

In the discussion above, we have assumed that the object $\mathbf{f}$ was variable, but we have not made use of its statistics in the design of the test statistic. If we know the scatter matrices of $\hat{\mathbf{f}}$, we can design a linear discriminant that will improve on either the simple matched filter or the prewhitening one. The general form of this discriminant is due to Hotelling,[9] though a special case was also considered by Fisher.[18] For binary decision problems, the Hotelling test statistic is

$$\lambda_{\mathrm{Hot}} = [\bar{\bar{\hat{\mathbf{f}}}}_2 - \bar{\bar{\hat{\mathbf{f}}}}_1]^t\hat{S}_2^{-1}\hat{\mathbf{f}}, \tag{76}$$

where the subscript Hot stands for Hotelling. The corresponding $\mathrm{SNR}^2$ can be derived from Eq. (65), with the result that

$$[\mathrm{SNR}_{\mathrm{Hot}}]^2 = \mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]. \tag{77}$$

This $\mathrm{SNR}^2$, sometimes called the Hotelling trace, is a common measure of class separability in pattern recognition.[10] Comparing Eq. (77) with Eq. (75), we see that the Hotelling SNR has the same structure in the case of general object variability as the prewhitening SNR does in the case of no object or signal variability.

## 5. RELATIONS BETWEEN ESTIMATION AND CLASSIFICATION METRICS

In the preceding two sections we have considered three specific estimators and three analogous test statistics for classification. The simple matched filter for classification is the counterpart of the simple ROI estimator, since each is computed simply by integrating $\hat{\mathbf{f}}$ over a template. Similarly, the prewhitening matched filter corresponds to the Gauss–Markov estimator, with both including a prewhitening step. Finally, the Wiener estimator and the Hotelling classifier both generalize the prewhitening operation and make explicit use of the second-order statistics of $\hat{\mathbf{f}}$.

These similarities in form suggest that there might be close relationships between the performances of the corresponding operators. On the other hand, the relationships are not obvious since different properties of the object $\mathbf{f}$ or its reconstruction $\hat{\mathbf{f}}$ are involved in estimation and classification. For example, the class separability as measured by the scatter matrices has no role in estimation, while bias is important in estimation but has no direct role in classification. (Bias plays an indirect role in classification, however, because system blur or artifacts can degrade classification performance, but that effect is accounted for in the scatter matrices.)

**Matched Filter versus Region-of-Interest Estimator**

In this section we derive and discuss the ratio of the performance metrics for a simple matched filter and the corresponding ROI estimator. Similar calculations for other test statistics and estimators are performed in subsequent sections. The ultimate goal is to determine to what extent a system optimized for a particular classification task is also optimal for a related estimation task.

The form that we choose for $\mathrm{SNR}_{\mathrm{MF}}$ is Eq. (69), which is based on the assumption that the correct template is used ($S = \hat{S}_1$). With $\mathrm{SNR}_{\mathrm{ROI}}$ from Eq. (47), the ratio of the metrics is

$$\frac{[\mathrm{SNR}_{\mathrm{MF}}]^2}{[\mathrm{SNR}_{\mathrm{ROI}}]^2} = \frac{\{\mathrm{tr}[\hat{S}_1]\}^2}{\mathrm{tr}[FW]} \frac{\mathrm{tr}[BW] + \mathrm{tr}[K_m W]}{\mathrm{tr}[\hat{S}_1 \hat{S}_2]}. \tag{78}$$

Multiplying by unity in various guises yields

$$\frac{[\mathrm{SNR}_{\mathrm{MF}}]^2}{[\mathrm{SNR}_{\mathrm{ROI}}]^2} = Q_{\mathrm{Rose}} Q_{\mathrm{bias}} Q_{\mathrm{conspic}} Q_{\mathrm{correl}}, \tag{79}$$

where

$$Q_{\mathrm{Rose}} = \frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}{\mathrm{tr}[FW]}, \tag{80}$$

$$Q_{\mathrm{bias}} = \frac{\mathrm{tr}[BW] + \mathrm{tr}[K_m W]}{\mathrm{tr}[K_m W]}, \tag{81}$$

$$Q_{\mathrm{conspic}} = \frac{\mathrm{tr}[K_m W]}{\mathrm{tr}[\hat{S}_2 W]}, \tag{82}$$

$$Q_{\mathrm{correl}} = \frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[\hat{S}_2 W]}{\mathrm{tr}[\hat{S}_1 \hat{S}_2]\mathrm{tr}[W]}. \tag{83}$$

We shall now discuss the interpretation of each of these factors in turn.

The reason for the designation of the Rose factor is most easily seen by considering the simplest of signal-detection tasks, where **s** is a small, low-contrast disk on a uniform, nonrandom background; the task is to decide whether the disk is present. If the disk is well resolved by the system, $\mathrm{tr}[\hat{S}_1]$ is essentially an integral of the square of the disk signal, or

$$\mathrm{tr}[\hat{S}_1] \simeq C^2 f_0{}^2 A_{\mathrm{disk}}, \tag{84}$$

where $f_0$ is the background level, $C$ is the contrast, and $A_{\mathrm{disk}}$ is the area of the disk measured in pixels. Similarly, if $W$ is a 0–1 matrix defining a region of interest, $\mathrm{tr}[W]$ is the area of that region in pixels, denoted $A_{\mathrm{ROI}}$, and $\mathrm{tr}[FW]$ is $f_0{}^2 A_{\mathrm{ROI}}$. Thus, for this simple problem,

$$Q_{\mathrm{Rose}} = C^2 A_{\mathrm{disk}}, \tag{85}$$

which is the dependence given by Rose's theory of disk detectability.[19] In more complicated detection problems, $\mathrm{tr}[FW]/\mathrm{tr}[W]$ is a reasonable measure of the background strength in the ROI, and $\mathrm{tr}[\hat{S}_1]$ is a reasonable measure of integrated average signal strength, so the Rose factor expresses, in a single scalar quantity, the effects of the size and contrast of the signal to be detected.

The interpretation of the second factor, $Q_{\mathrm{bias}}$, is straightforward. It is a factor that appears simply because bias is a component of the performance metric in estimation but not directly in classification. Note that $Q_{\mathrm{bias}}$ goes to unity if $B$ goes to zero.

The third factor is called the conspicuity factor, and it accounts for the excess noise in the ROI resulting from object variability or background clutter. If there is no object variability, $\hat{S}_2$ is equal to $K_m$, and $Q_{\mathrm{conspic}}$ goes to unity.

The last factor is called the correlation factor, and it approaches unity if the noise is uncorrelated and stationary, in which case $\hat{S}_2$ is a multiple of the unit matrix. It also approaches unity, regardless of the noise correlation, if the same integration template is used in the classification and estimation tasks ($\hat{S}_1 = W$). Thus $Q_{\mathrm{correl}}$ describes the different ways in which noise correlation affects classification and estimation tasks whenever the regions of integration are different. Note that $Q_{\mathrm{correl}}$ does not depend on the absolute noise level or signal contrast; if either $\hat{S}_1$ or $\hat{S}_2$ is multiplied by a constant, $Q_{\mathrm{correl}}$ is unchanged.

### Prewhitening Matched Filter versus Gauss–Markov Estimator

A somewhat more involved factorization occurs in the comparison of the prewhitening matched filter and the Gauss–Markov estimator. For the case where $S = \hat{S}_1$, the straightforward ratio of performance metrics from Eqs. (55) and (73) is

$$\frac{[\mathrm{SNR}_{\mathrm{PW}}]^2}{[\mathrm{SNR}_{\mathrm{GM}}]^2} = \frac{\{\mathrm{tr}[K_m{}^{-1}\hat{S}_1]\}^2}{\mathrm{tr}[FW]} \frac{\mathrm{tr}[(A^t K_m{}^{-1} A)^+ W]}{\mathrm{tr}[K_m{}^{-1}\hat{S}_2 K_m{}^{-1}\hat{S}_1]}. \tag{86}$$

We seek a factorization similar to that used in Eq. (79), but now there is no bias factor because the estimator is unbiased. This factorization is not unique, of course, but an interesting possibility is that

$$\frac{[\mathrm{SNR}_{\mathrm{PW}}]^2}{[\mathrm{SNR}_{\mathrm{GM}}]^2} = Q_{\mathrm{Rose}} Q_{\mathrm{conspic}} Q_{\mathrm{correl}}, \tag{87}$$

where the Rose factor is still given by Eq. (80) but

$$Q_{\mathrm{conspic}} = \frac{\mathrm{tr}[(A^t K_m{}^{-1} A)^+ W]}{\mathrm{tr}[\hat{S}_2 W]}, \tag{88}$$

$$Q_{\mathrm{correl}} = \frac{\{\mathrm{tr}[K_m{}^{-1}\hat{S}_1]\}^2 \,\mathrm{tr}[\hat{S}_2 W]}{\mathrm{tr}[K_m{}^{-1}\hat{S}_2 K_m{}^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}. \tag{89}$$

The conspicuity factor $Q_{\mathrm{conspic}}$ still measures the excess noise in the ROI resulting from object variability, but now the noise in the estimation task is amplified by the pseudoinverse operation. This factor now approaches unity only if $\hat{S}_2$ approaches $K_m$ and $A$ is well approximated by the unit matrix.

The correlation factor is more complicated. We first note that it is unchanged if any of the following matrices is multiplied by a scalar constant: $\hat{S}_1, \hat{S}_2, K_m, A$, or $W$. What it *does* depend on is the way in which noise correlation affects the two tasks. In spite of its complicated form, this factor still approaches unity for uncorrelated noise, where $\hat{S}_2$ and $K_m$ are proportional to the unit matrix.

Since ROI estimation is virtually the same operation as a nonprewhitening matched filter, we might expect a system optimized for ROI estimation also to be optimal for detection tasks based on the matched filter. That this is not necessarily so is shown by the four $Q$ factors. A change in the imaging system or processing algorithm might, for example, improve the contrast of the signal or its conspicuity without necessarily improving the estimation performance. Hanson[6] has found a case where a system optimized for classification performed rather poorly for estimation.

Table 1. Summary of All Comparisons

| Comparison | $[\mathrm{SNR}_{\mathrm{class}}]^2$ | $\frac{1}{[\mathrm{SNR}_{\mathrm{est}}]^2}$ | $Q_{\mathrm{Rose}}$ | $Q_{\mathrm{bias}}$ | $Q_{\mathrm{conspic}}$ | $Q_{\mathrm{correl}}$ |
|---|---|---|---|---|---|---|
| Matched filter versus ROI | $\frac{[\mathrm{tr}(\hat{S}_1)]^2}{\mathrm{tr}[\hat{S}_1\hat{S}_2]}$ | $\frac{\mathrm{tr}[BW]+\mathrm{tr}[K_mW]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[BW]+\mathrm{tr}[K_mW]}{\mathrm{tr}[K_mW]}$ | $\frac{\mathrm{tr}[K_mW]}{\mathrm{tr}[\hat{S}_2W]}$ | $\frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[\hat{S}_1\hat{S}_2]\mathrm{tr}[W]}$ |
| Prewhitening versus Gauss–Markov | $\frac{\{\mathrm{tr}[K_m^{-1}\hat{S}_1]\}^2}{\mathrm{tr}[K_m^{-1}\hat{S}_2K_m^{-1}\hat{S}_1]}$ | $\frac{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}{\mathrm{tr}[FW]}$ | 1 | $\frac{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}{\mathrm{tr}[\hat{S}_2W]}$ | $\frac{\{\mathrm{tr}[K_m^{-1}\hat{S}_1]\}^2\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[K_m^{-1}\hat{S}_2K_m^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}$ |
| Hotelling versus ROI | $\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]$ | $\frac{\mathrm{tr}[BW]+\mathrm{tr}[K_mW]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[BW]+\mathrm{tr}[K_mW]}{\mathrm{tr}[K_mW]}$ | $\frac{\mathrm{tr}[K_mW]}{\mathrm{tr}[\hat{S}_2W]}$ | $\frac{\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}$ |
| Hotelling versus Wiener | $\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]$ | $\frac{\mathrm{tr}[(A^tK_m^{-1}A+K_f^{-1})^{-1}W]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}{\mathrm{tr}[FW]}$ | $\frac{\mathrm{tr}[(A^tK_m^{-1}A+K_f^{-1})^{-1}W]}{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}$ | $\frac{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}{\mathrm{tr}[\hat{S}_2W]}$ | $\frac{\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}$ |

### Hotelling Discriminant versus Wiener Estimator

The factorization for comparison of the Hotelling classifier and the Wiener estimator again requires four factors since the estimator is biased. It uses Eq. (80) for the Rose factor and

$$Q_{\mathrm{bias}} = \frac{\mathrm{tr}[(A^tK_m^{-1}A + K_f^{-1})^{-1}W]}{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}, \tag{90}$$

$$Q_{\mathrm{conspic}} = \frac{\mathrm{tr}[(A^tK_m^{-1}A)^+W]}{\mathrm{tr}[\hat{S}_2W]}, \tag{91}$$

$$Q_{\mathrm{correl}} = \frac{\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[\hat{S}_1]\mathrm{tr}(W)}. \tag{92}$$

Note that the bias factor goes to unity if $K_f$ is large, which is the limit where the Wiener estimator approaches the unbiased Gauss–Markov estimator. In general, however, the Wiener estimator is biased toward the known ensemble mean of **f** (even though it is constructed so that the average bias over the whole class is zero). In contrast to the ROI estimator, the bias in the Wiener estimator serves to reduce the EMSE, so the bias factor is less than one.

The conspicuity factor here is the same as in the PW/GM comparison, while the correlation factor is the same as in the Hotelling/ROI comparison.

Little has been done experimentally on comparing performances of the Hotelling observer and Wiener estimators, but Smith and Barrett[20] did find a rough correlation between them in a study of pinhole coded apertures for gamma-ray imaging. Much further research on these comparisons is needed.

### Hotelling Discriminant versus Region-of-Interest Estimator

It might seem unfair to compare the optimum linear discriminant with the naïve ROI estimator, but in fact the more sophisticated estimators are only rarely used in practice, at least in medical imaging. In addition, there is some evidence, as mentioned above, that suggests that the Hotelling observer is often a good model for the human observer, and of course humans are the most common classifiers. Thus, in comparing the Hotelling observer and the ROI estimator, an argument can be made that we are simply comparing the methods actually used in practice.

Once again, we use Eq. (80) for the Rose factor, while the other factors are given by

$$Q_{\mathrm{bias}} = \frac{\mathrm{tr}[BW] + \mathrm{tr}[K_mW]}{\mathrm{tr}[K_mW]}, \tag{93}$$

$$Q_{\mathrm{conspic}} = \frac{\mathrm{tr}[K_mW]}{\mathrm{tr}[\hat{S}_2W]}, \tag{94}$$

$$Q_{\mathrm{correl}} = \frac{\mathrm{tr}[\hat{S}_2^{-1}\hat{S}_1]\mathrm{tr}[\hat{S}_2W]}{\mathrm{tr}[\hat{S}_1]\mathrm{tr}[W]}. \tag{95}$$

Thus the bias and conspicuity factors are the same as in matched filter versus ROI, while the correlation factor is the same as in Hotelling versus Wiener.

A summary of all the factors for all comparisons is given in Table 1.

## 6. DISCUSSION

### Human Observers

If the ultimate user of an image is a human observer, the figure of merit for image quality should have some relation to the performance of the human. Evaluation of the performance of a human observer by psychophysical studies and ROC analysis is routine[4] and will not be discussed in detail here. The objective of these studies, however, is usually to understand the characteristics of human perception, so the experiments are usually simple and easily analyzed. The signal to be detected in these experiments is often nonrandom and known to the observer; we refer to such experiments as signal-known-exactly (SKE). A consequence of the SKE assumption is that the ideal Bayesian observer uses a test statistic that is linear in the data. Since many SKE experiments show that the performance of a human observer tracks well with that of the Bayesian (with a statistical efficiency that can be attributed to internal noise in the human[21]), it is reasonable to assume that the human can perform some linear tasks rather well.

Even in the SKE case, however, human performance can be much inferior to Bayesian performance if the noise is strongly correlated.[22,23] One way to explain this observation is to say that the human cannot prewhiten the noise. A more straightforward explanation is based on the spatial-frequency-selective channels known to exist in the human visual system. Myers and Barrett[24] analyzed the performance of a Bayesian observer constrained to process the data through such channels. For a variety of SKE tasks, they found that the performance of this constrained or channelized Bayesian was indistinguishable from the performance of a nonprewhitening matched filter. Thus the channels provide an explanation of the apparent inability of a human to prewhiten.

The SKE investigations thus seem to suggest that a linear matched filter without prewhitening is a reasonable model for the human, at least in simple detection or discrimination tasks. It is not clear, however, what relevance SKE psychophysical studies have to more complex tasks such as medical diagnosis. As soon as the signal or background becomes random, the Bayesian must use a nonlinear function of the data as the test statistic. The Bayesian test statistic is the likelihood ratio (or log likelihood), which requires knowledge of the probability densities of the data under each hypothesis. It is virtually impossible to have such detailed information for realistic tasks, so we usually cannot calculate the performance of the Bayesian. In a few cases with modest degrees of signal uncertainty, where the nonlinear Bayesian was calculable, the human was found to be much inferior to the Bayesian, lending additional support to the use of linear models for assessment of image quality.[16]

Among the linear observer models, the Hotelling model has been successful in predicting human performance in several cases with object variability.[25,26] This model also explains human performance in any SKE task with uncorrelated noise since the nonprewhitening matched filter and the Hotelling observer are identical in these cases. The Hotelling observer does, however, involve a generalized prewhitening operation ($\hat{S}_2^{-1}$), so it still will not account for poor human performance in correlated noise. A natural extension of the Hotelling model would be to include the spatial-frequency channels, but that is beyond the scope of this paper.

### Implementation of the Hotelling Formalism

Our general scheme for implementing the Hotelling observer in medical applications is to begin with a realistic three-dimensional mathematical model of some organ or organ system, allowing variability both in normal anatomy and in the nature and placement of lesions or other pathology. This model is then used to create ensembles of objects in two or more classes (normal and abnormal classes in the simplest case), and these ensembles are used with an accurate model of the imaging system to create ensembles of images.

Given a training set of images created this way, a straightforward attempt to implement the Hotelling prescription would be to estimate the matrices $\hat{S}_1$ and $\hat{S}_2$ by sample covariance matrices and then try to form $\hat{S}_2^{-1}\hat{S}_1$ by usual matrix manipulations; this method fails badly. One problem is that the matrices are huge. If the original images are $64 \times 64$, then $\hat{S}_1$ and $\hat{S}_2$ are each $4096 \times 4096$. Furthermore, if the number of images in the training set is less than 4096, $\hat{S}_2^{-1}$ does not exist.

Several mathematical devices are needed to make the problem manageable. The first is to take advantage of the fact that we can create noise-free images, so that we can calculate $\hat{S}_2$ rigorously as the sum of two matrices, a noise-free part and a part resulting from Poisson noise [Eq. (33)]. The first part is estimated by sample covariance matrices formed from the training set, while the second part is computable if one knows the noise covariance matrix $K_n$, which is a diagonal matrix with diagonal elements given by the pixel averages of the objects in the training set. Since the second term in $\hat{S}_2$ is usually full rank and both terms are nonnegative definite, the inverse now exists, but it is still not practical to calculate it directly. Instead, we now take advantage of the fact that we do not need to know $\hat{S}_2^{-1}\hat{S}_1$ completely; rather, we need to find only its dominant eigenvectors and eigenvalues. Furthermore, we have already noted that $\hat{S}_1$ has rank $J-1$, where $J$ is the number of classes, so for a two-class problem we have to find only a single eigenvector and eigenvalue. The eigenvector is simply a $64 \times 64$ image in our example, so now we have a relatively routine image-reconstruction problem; an iterative algorithm for its solution is given by Fiete *et al.*[25] Once the eigenvector is found, its scalar product with each image generates a single scalar feature, which, however, contains all the information that was in the original image data set as far as this particular task is concerned. The Hotelling trace can then be readily calculated, and its variation with any number of engineering parameters can be efficiently studied.

### Estimation Tasks

For estimation tasks, the figure of merit should be based on the estimation algorithm actually used in practice, often just simple ROI estimation by integrating over a template. An estimate of the estimation SNR can then be obtained by generating a set of images, laying the template over each, and integrating. Since the actual object will be known for computer-generated images, the bias, variance, and SNR of the ROI estimate can be determined, with an accuracy dependent on the number of images.

The Wiener and Gauss–Markov estimators are more difficult to study since they require estimation of covariance matrices. One example of the use of the Wiener estimator for system optimization is the research of Smith and Barrett.[14]

The linear estimators discussed here do not, of course, exhaust the possibilities for using estimation tasks for system optimization. Mueller *et al.*[7] have recently investigated the use of nonlinear maximum-likelihood estimators for studying the performance of gamma-ray imaging systems in nuclear medicine. We do not know how the performance of these nonlinear estimators will track with the performance of linear estimators as some system parameter is varied.

**Relative Importance of Quantum Noise**

In quantum-limited imaging, we expect the SNR to vary as the square root of the number of quanta $N_q$ or the exposure time $T$, if all else is held fixed. The SNR's that we have calculated in this paper, however, also include other sources of randomness, so they will not necessarily have this simple dependence. To see what the dependence is, note that $\hat{S}_1$, being proportional to the square of the amplitude of the image, varies as $T^2$ if the processing algorithm is held constant as $T$ is varied. On the other hand $\hat{S}_2$ contains two terms, one describing the quantum noise and the other the object variability. Therefore it varies as $\alpha T + \beta T^2$.

In a classification problem with a fixed template (one that does not depend on the exposure time), the dependence of $\mathrm{SNR}^2$ on exposure time is thus given by [cf. Eq. (65)]

$$\mathrm{SNR}^2 \propto \frac{T^2}{\alpha T + \beta T^2}. \tag{96}$$

Thus $\mathrm{SNR}^2$ is proportional to $T$ at low exposure times, where the quantum noise dominates, but it saturates at large $T$, where the object variability limits the performance. Beyond this limit, which we call the conspicuity limit, there is no further advantage to collecting more quanta. This behavior has been observed experimentally in a number of studies in our group.[27,28]

If, on the other hand, the template is optimized at each exposure time, as in the Hotelling approach, $\mathrm{SNR}^2$ does not behave as in expression (96), and, in fact, it does not necessarily saturate.[29]

Similarly, in ROI estimation problems with a fixed template, $\mathrm{SNR}^2$ will initially increase as $T$ but then saturate as the bias dominates the statistical noise. Unbiased estimators such as Gauss–Markov will have an $\mathrm{SNR}^2$ that increases without bound. The Wiener estimator, on the other hand, displays a rather interesting behavior[14] with $T$, saturating at both large and small $T$. The Wiener SNR does not go to zero as $T$ goes to zero because we have some *a priori* information about the object and hence about the parameter $\theta$ being estimated. The SNR will also saturate at large $T$, because of bias, unless $\theta$ satisfies the estimability condition [Eq. (38)].

**Choice of Task for System Optimization**

The designer of an imaging system will want to select one figure of merit with which to optimize the system. If the system is intended for a single, specific task, the results in this paper provide a general matrix expression for a suitable figure of merit. As we have demonstrated in a number of other publications,[14,20,25–29] these figures of merit can be numerically evaluated and used for system optimization. The main concern here is to use realistic models for the task, the object class, and the imaging system. If the task is not realistic, the resulting system may, in fact, perform rather poorly.[29]

A major difficulty arises, however, when the system is to be used for many tasks, perhaps some estimation and some classification tasks. A key question is then the following: To what extent is a system that is optimized for estimation also optimized for classification? The factorization formulas provide a framework for answering this question for particular systems. Any change to the system that does not alter any of the $Q$ factors will affect estimation and detection performance in the same way, but many examples of changes that do alter one or more $Q$ factors are easily found. For example, if the signal to be detected is small and barely resolvable, the Rose factor is sensitive to any change that affects the spatial resolution of the system and hence the contrast of the signal. Limited spatial resolution is also a form of bias in estimation if the spatial extent of $\mathbf{w}$ is small, but other forms of image artifact also contribute to the bias factor. The conspicuity factor may also be affected by image blur, but it matters greatly where the blurring occurs. Blurring by the image-forming elements (the so-called aperture blur in the terminology of Wagner *et al.*) will reduce $\mathrm{tr}[\hat{S}_2 W]$ but leave $\mathrm{tr}[K_m W]$ almost unchanged, since $\mathbf{m}$ is white noise if there is no postprocessing of the image. Postdetection image smoothing, on the other hand, will reduce both $\mathrm{tr}[\hat{S}_2 W]$ and $\mathrm{tr}[K_m W]$ and might leave the conspicuity factor relatively unchanged. The correlation factor is insensitive to aperture blur since it equals one for white noise, and it is probably relatively insensitive to postdetection processing as well.

When the system is to be used for a variety of tasks, the SNR's for all relevant tasks should be evaluated, but the $Q$ factors should be studied as well. In this way the designer can have some confidence that the system is not being unacceptably degraded for, say, an estimation task when it is being optimized for a classification task.

## 7. CONCLUSIONS

We have presented a general formalism under which the performance of linear observers can be calculated for classification or estimation tasks. Three specific observers were considered in the classification case: the simple matched filter, the prewhitening matched filter, and the Hotelling observer. Similarly, three linear estimators were considered: simple ROI integration, Gauss–Markov, and Wiener. In each case the performance was expressed as a SNR. These expressions include the effects of both quantum noise and object variability.

We then examined the relationships between SNR's for classification and estimation. We wrote the classification SNR as an estimation SNR times four factors. These factors account for the effects of signal size and contrast in detection or classification, the effects of bias in estimation, the effects of signal conspicuity, and the effects of noise correlation. In principle, these factors provide a framework for understanding how various changes in the imaging system or the processing affect classification and estimation performance, but much further research is needed to evaluate and understand them in particular cases.

Further research is also needed to clarify the effects of the linearity assumptions. We need to know whether a human observer can perform nonlinear operations, and, if so, we need to incorporate that information into figures of merit for imaging systems whenever the ultimate user is a human.

In spite of these uncertainties, the author believes that the figures of merit presented here are valid and useful tools for the optimization of imaging systems. All these figures of merit relate to specific, realistic tasks that the system and observer must perform, and they can all be evaluated numerically, though sometimes with great expenditure of computer time since they all require simulation of an ensemble of images. This computational expense may simply be unavoidable if the vague concept of image quality is ever to have a precise meaning.

## APPENDIX A: STATISTICS OF THE REGION-OF-INTEREST ESTIMATOR

From Eqs. (19) and (40), we have

$$\hat{\theta}_{\mathrm{ROI}} = \mathbf{w}^t\hat{\mathbf{f}} = \mathbf{w}^t[\mathbf{f} + \mathbf{b} + \mathbf{m}]. \tag{A1}$$

The variance is given by

$$\begin{aligned}\mathrm{var}(\hat{\theta}_{\mathrm{ROI}}) &\equiv \langle(\hat{\theta}_{\mathrm{ROI}} - \bar{\bar{\hat{\theta}}}_{\mathrm{ROI}})^2\rangle_{n,f} \\ &= \langle[\mathbf{w}^t(\hat{\mathbf{f}} - \bar{\bar{\hat{\mathbf{f}}}})]^2\rangle_{n,f} \\ &= \mathbf{w}^t\langle(\hat{\mathbf{f}} - \bar{\bar{\hat{\mathbf{f}}}})(\hat{\mathbf{f}} - \bar{\bar{\hat{\mathbf{f}}}})^t\rangle_{n,f}\mathbf{w} \\ &= \mathbf{w}^t\hat{K}\mathbf{w} = \mathrm{tr}[\hat{K}W],\end{aligned} \tag{A2}$$

where the definition of $\hat{K}$ from Eq. (25) has been used. Equation (A2) is the same as Eq. (44) in Section 3.

To calculate the EMSE, note that

$$\hat{\theta}_{\mathrm{ROI}} - \theta = \mathbf{w}^t\mathbf{b} + \mathbf{w}^t\mathbf{m}, \tag{A3}$$

so that

$$\begin{aligned}\mathrm{EMSE}[\hat{\theta}_{\mathrm{ROI}}] &= \langle(\hat{\theta}_{\mathrm{ROI}} - \theta)^2\rangle_{n,f} \\ &= \mathbf{w}^t\langle\mathbf{b}\mathbf{b}^t\rangle_{n,f}\mathbf{w} + \mathbf{w}^t\langle\mathbf{m}\mathbf{m}^t\rangle_{n,f}\mathbf{w} \\ &= \mathbf{w}^tB\mathbf{w} + \mathbf{w}^tK_m\mathbf{w} = \mathrm{tr}(BW) + \mathrm{tr}(K_mW),\end{aligned} \tag{A4}$$

where we assume that $\mathbf{m}$ has zero mean and is uncorrelated with $\mathbf{b}$. Equation (A4) agrees with Eq. (45) in the text.

## APPENDIX B: STATISTICS OF THE GAUSS-MARKOV ESTIMATOR

Derivation of the EMSE for the Gauss–Markov estimator will proceed more smoothly if we first establish that it is indeed an unbiased estimator. This proof is nontrivial, requiring some properties of the pseudoinverse matrix.

From Eqs. (34) and (53),

$$\hat{\theta}_{\mathrm{GM}} - \theta = \mathbf{v}^t\hat{\mathbf{f}} - \mathbf{w}^t\mathbf{f}, \tag{B1}$$

where

$$\mathbf{v}^t = \mathbf{w}^t[K_m^{-1/2}A]^+K_m^{-1/2}. \tag{B2}$$

The conditional mean of $\hat{\theta}_{\mathrm{GM}}$ is

$$\begin{aligned}\langle\hat{\theta}_{\mathrm{GM}}\rangle_{n|f} &= \mathbf{w}^t[K_m^{-1/2}A]^+K_m^{-1/2}A\mathbf{f} \\ &= \mathbf{w}^tA^+A[K_m^{-1/2}A]^+K_m^{-1/2}A\mathbf{f},\end{aligned} \tag{B3}$$

where we have used Eq. (18) in the first step and the estimability condition [Eq. (38)] in the second. [Note that $A^+A$ is symmetric, so that Eq. (38) is equivalent to $\mathbf{w}^tA^+A = \mathbf{w}^t$.]

We can insert a unit matrix, in the form $K_m^{1/2}K_m^{-1/2}$, so that Eq. (B3) becomes

$$\langle\hat{\theta}_{\mathrm{GM}}\rangle_{n|f} = \mathbf{w}^tA^+K_m^{1/2}K_m^{-1/2}A[K_m^{-1/2}A]^+K_m^{-1/2}A\mathbf{f}. \tag{B4}$$

We now use the Penrose equation,[8] which says that, for any matrix $D$,

$$DD^+D = D. \tag{B5}$$

If we take $D = K_m^{-1/2}A$, Eq. (B4) becomes

$$\langle\hat{\theta}_{\mathrm{GM}}\rangle_{n|f} = \mathbf{w}^tA^+A\mathbf{f} = \mathbf{w}^t\mathbf{f} = \theta, \tag{B6}$$

where we have again used the estimability condition. Equation (B6) shows that $\hat{\theta}_{\mathrm{GM}}$ is indeed unbiased.

The derivation of the EMSE is now straightforward. We write

$$\begin{aligned}\mathrm{EMSE}[\hat{\theta}_{\mathrm{GM}}] &= \langle[\hat{\theta}_{\mathrm{GM}}]^2\rangle_{n,f} - \langle\theta^2\rangle_{n,f} \\ &= \mathbf{v}^t\langle\hat{\mathbf{f}}\hat{\mathbf{f}}^t\rangle_{n,f}\mathbf{v} - \mathbf{w}^t\langle\mathbf{f}\mathbf{f}^t\rangle_f\mathbf{w} \\ &= \mathbf{v}^t[AFA^t]\mathbf{v} + \mathbf{v}^tK_m\mathbf{v} - \mathbf{w}^tF\mathbf{w},\end{aligned} \tag{B7}$$

where we have used the definition of $F$ from Eq. (48). We now use the fact that $\hat{\theta}_{\mathrm{GM}}$ is unbiased, which means that $\mathbf{v}^tA\mathbf{f} = \mathbf{w}^t\mathbf{f}$ and hence $\mathbf{v}^t[AFA^t]\mathbf{v} = \mathbf{w}^tF\mathbf{w}$. Thus

$$\mathrm{EMSE}[\hat{\theta}_{\mathrm{GM}}] = \mathbf{v}^tK_m\mathbf{v} = \mathrm{tr}[VK_m] = \mathrm{tr}[(A^tK_m^{-1}A)^+W], \tag{B8}$$

which is in agreement with Eq. (54).

## ACKNOWLEDGMENTS

Much of the impetus for this research came from an ongoing collaboration with Robert Wagner and Kyle Myers of the Center for Devices and Radiological Health, U.S. Food and Drug Administration. The author is deeply grateful for their advice and helpful comments throughout the course of the research. Stimulating discussions with Ted Gooley, Jannick Rolland, William Dallas, Gabor Herman, and David Brown and the assistance of John Aarsvold and Robert Rowe in carefully reviewing the manuscript are also gratefully acknowledged. This work was supported by the National Cancer Institute under grant P01-CA23417.

## REFERENCES

1. L. B. Lusted, "Signal detectability and medical decision-making," Science **171,** 1217–1219 (1971).
2. C. F. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. 21, 720–733 (1986).
3. J. A. Swets, "Measuring the accuracy of diagnostic systems," Science **240,** 1285–1293 (1988).
4. J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems* (Academic, New York, 1982).
5. K. M. Hanson, "POPART—performance optimized algebraic reconstruction technique," in *Visual Communications and Image Processing III*, T. R. Hsing, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1001,** 318–325 (1989).

6. 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 H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090**, 146–153 (1989); "Method of evaluating image-recovery algorithms based on task performance," J. Opt. Soc. A **7**, 1294–1304 (1990).
7. 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. (to be published).
8. A. Albert, *Regression and the Moore–Penrose Pseudoinverse* (Academic, New York, 1972).
9. H. Hotelling, "The generalization of Student's ratio," Ann. Math. Stat. **2**, 360–378 (1931).
10. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972).
11. H. C. Andrews and B. R. Hunt, *Digital Image Restoration* (Prentice-Hall, Englewood Cliffs, N.J., 1977).
12. H. H. Barrett and W. Swindell, *Radiological Imaging: Theory of Image Formation, Detection and Processing* (Academic, New York, 1981), Vols. I and II.
13. J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory* (McGraw-Hill, New York, 1978).
14. W. E. Smith and H. H. Barrett, "Linear estimation theory applied to the evaluation of *a priori* information and system optimization in coded-aperture imaging," J. Opt. Soc. Am. A **5**, 315–330 (1988).
15. H. L. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).
16. R. K. 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 H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090**, 183–194 (1989).
17. R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," Phys. Med. Biol. **30**, 489–518 (1985).
18. T. Y. Young and T. W. Calvert, *Classification, Estimation and Pattern Recognition* (Elsevier, New York, 1974).
19. A. Rose, "The sensitivity performance of the human eye on an absolute scale," J. Opt. Soc. Am. **38**, 196–208 (1948).
20. 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).
21. A. E. Burgess and B. Colbourne, "Visual signal detection. IV. Observer inconsistency," J. Opt. Soc. Am. A **5**, 617–627 (1988).
22. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," Science **214**, 93–94 (1981).
23. 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).
24. 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).
25. 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).
26. T. A. White, H. H. Barrett, E. B. Cargill, R. D. Fiete, and M. Ker, "The use of the Hotelling trace to optimize collimator performance," J. Nucl. Med. **30**, 892(A) (1989).
27. D. P. Kwo, H. B. Barber, H. H. Barrett, T. S. Hickernell, and J. M. Woolfenden, "Comparison of NaI(T1), $HgI_2$ and CdTe surgical probes—II: Effect of scatter compensation on probe performance," submitted to Med. Phys.
28. T. S. Hickernell, H. H. Barrett, H. B. Barber, J. N. Hall, and J. M. Woolfenden, "Probability modelling of a radiation-detector-probe system for statistical analysis," Phys. Med. Biol. (to be published).
29. K. J. Myers, J. P. Rolland, H. H. Barrett, and R. F. Wagner, "Effect of a spatially varying background on aperture choice for optimized image quality," J. Opt. Soc. Am. A **7**, 1279–1293 (1990).