Published in final edited form as:

J Opt Soc Am A Opt Image Sci Vis. 2015 April 1; 32(4): 549–565.

Method for optimizing channelized quadratic observers for binary classification of large-dimensional image datasets

M. K. Kupinski^{1,2,*} and E. Clarkson^{1,2}

¹College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA

²Department of Medical Imaging, University of Arizona, Tucson, Arizona 85721, USA

Abstract

We present a new method for computing optimized channels for channelized quadratic observers (CQO) that is feasible for high-dimensional image data. The method for calculating channels is applicable in general and optimal for Gaussian distributed image data. Gradient-based algorithms for determining the channels are presented for five different information-based figures of merit (FOMs). Analytic solutions for the optimum channels for each of the five FOMs are derived for the case of equal mean data for both classes. The optimum channels for three of the FOMs under the equal mean condition are shown to be the same. This result is critical since some of the FOMs are much easier to compute. Implementing the CQO requires a set of channels and the first- and second-order statistics of channelized image data from both classes. The dimensionality reduction from M measurements to L channels is a critical advantage of CQO since estimating image statistics from channelized data requires smaller sample sizes and inverting a smaller covariance matrix is easier. In a simulation study we compare the performance of ideal and Hotelling observers to CQO. The optimal CQO channels are calculated using both eigenanalysis and a new gradient-based algorithm for maximizing Jeffrey's divergence (J). Optimal channel selection without eigenanalysis makes the J-CQO on large-dimensional image data feasible.

1. INTRODUCTION

Our work is motivated by a challenge that is common in many imaging applications: sorting image data between two classes of objects (e.g., signal present and signal absent) when linear classifiers do not perform well enough for the application. An optimal quadratic classifier requires either a training set of images from each class or prior knowledge of the first- and second-order statistics of the image data from each class. The first-order statistics are the average images from each class and the second-order statistics are the covariance matrices from each class. If a training set of images is available the first- and second-order sample statistics can be used. Optimal quadratic classifiers are difficult to compute in imaging applications because of the large number of measurements made by most imaging systems. A single image can contain a few million elements and the number of elements in the covariance matrix is equal to the square of this number. When working with the covariance matrix, storing it can be challenging, inverting it can be impractical, and

*Corresponding author: meredith@optics.arizona.edu.

accurately estimating it from finite training data can even be impossible. Our work addresses this big data problem by using a quadratic classifier on images that have been reduced in size by a linear transformation; we will refer to this as a channelized quadratic observer (CQO). This approach demands answering the following question: which linear transform is best for computing a quadratic classifier for a given imaging application? To address this question we have developed a new method for optimizing CQOs for binary classification of large-dimensional image datasets.

To introduce the detection method, begin by considering the relationship between an image and an object as

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

Here, $\mathbf{g}$ is an $M \times 1$ vector of measurements made by an imaging system that is represented as a continuous-to-discrete operator $\mathcal{H}$; the measurements of the continuous object f are corrupted by measurement noise $\mathbf{n}$. We will consider post-processing signal detection. That is to say, the forward imaging model $\mathcal{H}$ is fixed and can even be unknown since only the statistics of the image data will be used. We are interested in linear combinations of the image data of the form

$$\mathbf{v} = \mathbf{T}\mathbf{g}, \quad (2)$$

where $\mathbf{T}$ is an $L \times M$ matrix and compression is achieved since $L < M$. Using terminology from the perception literature, each row of $\mathbf{T}$ is referred to as a channel [1]. In this paper, $\mathbf{v}$ will be called a channelized image or a channelized data vector. Mathematical observers for detection or classification tasks operate on image data $\mathbf{g}$, or preferably channelized image data $\mathbf{v}$, to form a scalar-valued decision variable. Objective assessment of image quality [2] quantifies the ability of an observer to use image data for performing the scientific task of interest, e.g., detection, classification, or estimation. Channelized data are preferable for mathematical observers since calculating a decision variable usually involves the estimation of parametric likelihoods [3]. In the channelized representation this estimation can be much more accurate given common constraints on finite training data [4,5]. Computational needs are lower because the inverse of a covariance matrix, required for likelihood evaluations, is now $L \times L$ instead of $M \times M$.

In this work we will present gradient-based optimization methods for finding the solution to $\mathbf{T}$ (once L is selected) that maximizes detection task performance of the ideal observer (i.e., the likelihood ratio) given Gaussian statistics on the channel outputs, $\mathbf{v}$, for both classes. We will consider the first- and second-order statistics of each class to be different, in general, which leads to a quadratic relationship between the likelihood ratio and the image data; we call this a quadratic observer. When the second-order statistics are equal the ideal observer is linear and the optimal solution for $\mathbf{T}$ is the Hotelling observer (i.e., a prewhitened match filter). This equal covariance assumption is valid when the two classes differ by the addition of a signal that is weak enough, relative to other sources of variability, so that it does not affect the covariance matrix. When the means are equal but the covariances are different we show a new result: that the same optimal $\mathbf{T}$ solution is achieved using optimization with

respect to the Bhattacharyya distance, Jeffrey's divergence, and the area under the curve (AUC). This equal mean assumption is valid in ultrasound imaging [6–8] and in many texture discrimination tasks.

The next section is devoted to a review of related work. Assumptions and notation are established in Section 3. Then we show an analytic gradient, with respect to the linear channels, for the following: Section 4) Kullback–Liebler (KL) divergence [9]; Section 5) the symmetrized KL divergence (also referred to as Jeffrey's divergence (J) in information theory [10]); Section 6) the Bhattacharyya distance [11] (also called $G(0)$ in [12]); and Section 7) the area under the ideal-observer receiver operating characteristic (ROC) curve, also known as the AUC [13,14]. We will see by the end of these Sections that the J and $G(0)$ metrics are maximized at the same set of channels that maximizes the AUC when there is no signal in the mean. This results in an important surrogate figure of merit (SFOM) [15] for imaging applications since J is much easier to compute than the AUC.

In Section 8 we focus on the specific case of no signal in the mean and compute the Hessian of J, with L number of linear channels. We use this Hessian to show that there are no more than $L + 1$ local maxima in this case, one of which is the global maximum. This is useful information for a gradient-based algorithm to find the global maximum.

In Section 9 the results of a simulation study are presented. Here, the AUC is computed for the ideal and Hotelling observers and compared to CQO performance. The CQO is computed in two different ways: 1) an iterative algorithm based on a gradient search to maximize J, and 2) an observer introduced in Section 5 that is based on an eigenanalysis of the covariance matrices and is optimal when there is no signal in the mean. The similar task performance between the J-based CQO (J-CQO) and the much less tractable Eigen-CQO indicates strong potential for the J-CQO method in more realistic imaging applications.

2. RELATED WORK

A. Channels in Imaging

For medical imaging applications image channels have been used to approximate Hotelling observers with channelized Hotelling observers (CHO) [16], both of which are linear observers. Channels have also been used to approximate the ideal observer, which is not necessarily linear in the image data [17,18]. These channelized ideal observers (CIOs) have been explored for both standard channels used for CHOs and channels derived from the imaging system operator [19,20].

In computer vision the interpretation of a single image is decomposed into subimage detection and classification tasks that are customized to the desired data product. Here, an image channel can be related to the original image data by both linear and nonlinear transforms; even the channel outputs themselves can be combined to further reduce the original data to features. A succinct review of different types of image channels and feature selection methods in this community is provided in [21].

Across imaging applications, the selection of channels can be motivated by maximizing the performance of the channelized observer; the channels are then referred to as efficient

channels [2]. On the other hand anthropomorphic channels are designed to approximate human observer performance and are often based on properties of the human visual system [22–24]. The *kernel trick* used for support vector machines (SVM) employs a nonlinear function of the data and a linear discriminant on that function's output [25]. SVM seeks nonlinear channels, either efficient or anthropomorphic depending on the application, for a linear observer. This work concentrates on efficient linear channels for quadratic observers.

B. Eigenanalysis for Compression

In this paper, we consider the task of detection among two classes and study the eigenanalysis of the two covariance matrices $\mathbf{K}_1$ and $\mathbf{K}_2$. When these covariance matrices are unequal the data is called heteroscedastic. Covariance matrix eigenanalysis is rarely practical for modern imaging systems since an image is comprised of several million elements. We will show a new and computational feasible approach for optimizing channel selection, which we denote J-CQO, and compare its detection task performance with an eigenanalysis approach, which we denote Eigen-CQO. We will also show that Eigen-CQO is optimal when the data is heteroscedastic and the mean images are equal. In [26] Fukunaga and Koontz were the first to suggest covariance matrix eigendecomposition for detection and classification tasks; this approach is called the Fukunaga and Koontz transform (FKT). The FKT uses a matrix $\mathbf{P}$ to transform the data so that $\mathbf{P}(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{P}^\dagger$ equals the identity matrix. This equality guarantees that both covariance matrices of the transformed data will have the same eigenvectors. Furthermore, the sum of the two eigenspectra, when eigenvalues associated with the same eigenvector are added, is equal to one. Consequently, the transformation by $\mathbf{P}$ makes the *strongest* eigenvectors of one class the *weakest* eigenvectors of the other. In [27] the statistical properties of the FKT are studied and it is reported that, with certain eigenspectrum assumptions, the FKT is the optimal low-rank approximation to the optimal classifier for zero mean, heteroscedastic, and normally distributed data. FKT is widely used in pattern recognition; its adaptation is called a tuned basis function (TBF) in [28] for finding an optimal solution to $\tilde{\mathbf{T}}$ to use in the quadratic test statistic, $\mathbf{g}^\dagger \tilde{\mathbf{T}} \mathbf{g}$.

In Section 5, we will show that with the normality assumptions used in this paper, and when there is no signal in the mean, an optimal channel matrix for the J FOM can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l .

These eigenvectors are chosen to have the L largest values of $\kappa_l + \kappa_l^{-1}$ and we call this observer the Eigen-CQO. Using these same assumptions on the statistics of the data, it was shown in [29] that the Eigen-CQO maximizes the Bhattacharya distance between two classes. Multiple discriminant analysis (MDA), described in [30], is similar to Eigen-CQO but instead of $\mathbf{K}_1$ and $\mathbf{K}_2$ an intraclass and an extraclass scatter matrix are estimated from finite data; the imaging task is to distinguish among more than two classes. MDA is achieved when the scatter matrices are estimated from N image samples, the FKT eigenspectrum is calculated from a QR decomposition on an $N \times M$ matrix, and the final classification decision is formed using a 1-nearest neighbor (1-NN) algorithm.

C. Information Metrics in Imaging

A SFOM is a quantity that correlates with ideal-observer task performance, as measured by the AUC or some other task-based figure of merit (FOM), but that is easier to compute [31]. Statistical distances and divergences used in information theory can be powerful SFOM in imaging if they are proven, analytically or empirically, to correlate with task performance. In [32] it is recognized that J is an alternative FOM to linear discriminant analysis (LDA), which is needed when second-order statistics are unequal among the classes. However, instead of optimizing J directly an upper-bound that is quadratic in the channels is used. Once a channel solution is obtained the classification decision is formed using a 1-NN algorithm and accuracy is reported, as opposed to the channelized ideal observer and ROC analysis used in this work. The KL-divergence and Chernoff distances were used in [33] to quantify the effect of compression when the matrix $\mathbf{T}$ is populated with random entries and the statistics are described by uncorrelated normal distributions. These results describe upper- and lower-bounds for the J divergence, which are valid with high probability for random channel matrices. In this paper, we are interested in the exceptions to these bounds, i.e., the solutions to $\mathbf{T}$ that produce large values of J and are unlikely samples from a random distribution.

D. Relation to Compressive Sensing

In the unrealistic limit of infinite computational resources, and perfect knowledge of the image statistics, efficient channels for mathematical observers are not necessary. As pointed out in [34] “Conventional sensing systems typically first acquire data in an uncompressed form (e.g., individual pixels in an image) and then perform compression for storage or transmission. In contrast, compressive sensing (CS) researchers would like to acquire data in an already compressed form, reducing the quantity of data that need be measured in the first place.” This work on channelized observers is a post-processing method that is complementary to CS. Given the large popularity of this area of research we will remark on the relationship to CS for the sake of clarity and context.

For CS we expand the object function as a finite series in terms of basis functions

$$f(\mathbf{r}) = \sum_{n=0}^N f_n \psi_n(\mathbf{r}). \quad (3)$$

The imaging equation [see Eq. (1)] is then replaced with

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

The $M \times N$ matrix $\mathbf{H}$ is called the sensing matrix or system matrix. The astonishing breakthrough of CS was based on optimal signal recovery by selecting $\mathbf{H}$ to take advantage of sparsity constraints on $\mathbf{f}$ [35]. This leads to asymptotic relationships between residuals of the recovered signal and stochastic models for $\mathbf{H}$, $\mathbf{f}$, and $\mathbf{n}$ [36]. However, even in some of the early CS studies, it was recognized that improved signal recovery could be achieved when knowledge of the underlying signal of interest was used to select $\mathbf{H}$ [37]. Research in this field has expanded to include evaluating candidate solutions to $\mathbf{H}$ with respect to

information-based metrics from communications theory, such as Shannon information [38]. It was pointed out in [39] that an information-theoretic metric is particularly attractive as it allows an upper-bound on the task-specific performance of an imager, independent of the algorithm used to extract the relevant information. Although this might be advantageous for hardware design, for post-processing of the image data an algorithm, i.e., mathematical observer, is required. Metrics related to estimation, classification, and detection tasks have also been investigated for CS applications; these metrics are then a function of both $\mathbf{H}$ and the mathematical observer [40]. To compare multiple solutions for $\mathbf{H}$ a constraint must be used, such as equivalent number of photon counts, so that the noise statistics are part of the trade-off with the measurement scheme and the compression ratio. Increasing the quantity of measurements of $\mathbf{f}$ naturally transmits more information about the object, but this increase in the dimensionality of $\mathbf{H}$ is penalized by a larger measurement noise contribution. The extreme case of a single measurement of $\mathbf{f}$ integrates all available photons and suffers minimal measurement deviation due to noise. The optimal solution for $\mathbf{H}$ will depend upon the FOM, the statistics of $\mathbf{f}$, and the statistics of $\mathbf{n}$. By contrast, this work concentrates exclusively on post-processing for detection; a test statistic is calculated from the image data via Eq. (9). We seek the optimal solution for $\mathbf{T}$, which depends upon the chosen FOM for the detection task and the statistics of $\mathbf{g}$. An important distinction between these two problems is that in post-processing $\mathbf{T}$ is applied to $\mathbf{H}\mathbf{f}$ and to $\mathbf{n}$. Thus, the noise statistics of the channelized image data depends on the channel matrix $\mathbf{T}$.

3. FORMULATION OF THE PROBLEM

As a first step, we will introduce some notation and describe the problem that we are considering. All vectors are column vectors unless otherwise specified. The M -dimensional vector $\mathbf{g}$ will represent the input to the classifier, which will classify this vector as either belonging to the population corresponding to the probability density function (PDF) $pr_1(\mathbf{g})$ or the population corresponding to the PDF $pr_2(\mathbf{g})$. This vector may be a direct image, a reconstructed image, or the raw data being produced by an imaging system.

The ideal classifier uses the log-likelihood ratio

$$\lambda_I(\mathbf{g}) = \ln[pr_1(\mathbf{g})] - \ln[pr_2(\mathbf{g})], \quad (5)$$

as a decision variable and compares the result to a threshold. If the decision variable is above the threshold, then the data vector is assigned to $pr_1(\mathbf{g})$, and otherwise it is assigned to $pr_2(\mathbf{g})$. This observer maximizes the AUC, as well as other task-based FOMs [2,14]. In imaging applications the dimension M of the vector $\mathbf{g}$ is very large. We will assume for $i = 1, 2$ that $pr_i(\mathbf{g})$ is a PDF with mean $\bar{\mathbf{g}}_i$ and covariance matrix $\mathbf{K}_i$. This creates two problems when we try to implement the ideal observer for the classification task. The first problem is computational; even for Gaussian PDFs we will have to invert two $M \times M$ covariance matrices $\mathbf{K}_i$, which may not be feasible if, for example, the input images contain millions of pixels. The second problem is that, if we are estimating the image statistics from data, which is often the case, we will need a very large number of samples to get reliable estimates. For example, in the Gaussian case, the number of samples needs to be at least M to get invertible estimates of the $\mathbf{K}_i$, and typically needs to be an order of magnitude greater to get reliable

estimates. This provides a motivation for trying to reduce the dimension of the data vector before implementing the ideal observer.

The data reduction will be implemented by an $L \times M$ dimensional matrix $\mathbf{T}$ via the equation $\mathbf{v} = \mathbf{T}\mathbf{g}$. We refer to the vector $\mathbf{v}$ as the channelized data vector and the rows of $\mathbf{T}$ as the channels for the data reduction. The number L is the dimension of the channelized data and satisfies $L \ll M$. We will always assume that $\mathbf{T}$ is a full rank matrix so that the channels are linearly independent. We will assume that the PDFs for the channelized data for the two populations are given by normal distributions

$$pr_i(\mathbf{v}) = \left[(2\pi)^L \det(\mathbf{TK}_i\mathbf{T}^\dagger) \right]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_i \right)^\dagger (\mathbf{TK}_i\mathbf{T}^\dagger)^{-1} \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_i \right) \right], \quad (6)$$

for $i = 1, 2$. The channelized data may be at least approximately Gaussian distributed even if the original data are not. In practice, almost all useful channel matrices are fully populated. Therefore, if the original data components are sufficiently independent from each other, then we can argue from the central limit theorem that the channel outputs will be approximately Gaussian distributed for almost all channel matrices [41]. In this case, we would expect the methods developed in this work to produce channel matrices that are very near optimal in terms of the FOMs discussed below.

Given a fixed set of L channels the optimal classifier will compute the channelized log-likelihood

$$\lambda(\mathbf{T}, \mathbf{g}) = \ln[pr_1(\mathbf{T}\mathbf{g})] - \ln[pr_2(\mathbf{T}\mathbf{g})], \quad (7)$$

and compare this decision variable to a threshold. For computational purposes we define, for $i = 1, 2$

$$Q_i(\mathbf{v}) = \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_i \right)^\dagger (\mathbf{TK}_i\mathbf{T}^\dagger)^{-1} \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_i \right). \quad (8)$$

Then we have

$$\lambda_{CQO}(\mathbf{T}, \mathbf{g}) = \frac{1}{2} \left[Q_2(\mathbf{T}\mathbf{g}) - Q_1(\mathbf{T}\mathbf{g}) - \ln \det(\mathbf{TK}_1\mathbf{T}^\dagger) + \ln \det(\mathbf{TK}_2\mathbf{T}^\dagger) \right], \quad (9)$$

where the notation denotes the CQO. The last two terms are usually dropped, since they do not depend on the data. We will not be doing this because we want to vary the channel matrix to search for an optimal set of channels for a given L and therefore we need an unmodified likelihood ratio. All of the FOMs we will be considering depend on moments of $\lambda_{CQO}(\mathbf{T}, \mathbf{g})$, which we will be referring to as $\lambda(\mathbf{v})$.

For any function $f(\mathbf{v})$ we will use the notation $\langle f(\mathbf{v}) \rangle_i$ for the expectation of $f(\mathbf{v})$ in each class,

$$\langle f(\mathbf{v}) \rangle_i = \int_{\mathbb{R}^L} f(\mathbf{v}) pr_i(\mathbf{v}) d^L v. \quad (10)$$

The following FOMs will be examined and optimal channels will be determined for each. The FOMs are listed in increasing order of the difficulty of the channel optimization calculations. The first two FOMs are Kullback–Liebler (KL) divergences,

$$F_1(\mathbf{T}) = \langle \lambda(\mathbf{v}) \rangle_1 = D_{KL}(pr_1 \| pr_2), \quad (11)$$

$$F_2(\mathbf{T}) = -\langle \lambda(\mathbf{v}) \rangle_2 = D_{KL}(pr_2 \| pr_1); \quad (12)$$

we will refer to these FOMs as KL1 and KL2, respectively. (Note that, from here onward, pr_i will refer to the corresponding PDF for the channelized data.) These two FOMs are measures of the difference between two PDFs and are used extensively in information theory. The third FOM is the symmetrized KL divergence, commonly referred to as J,

$$F_3(\mathbf{T}) = \langle \lambda(\mathbf{v}) \rangle_1 - \langle \lambda(\mathbf{v}) \rangle_2. \quad (13)$$

This symmetric measure of the difference between two PDFs has been related to the ideal-observer AUC in applications where $\mathbf{g}_1 = \mathbf{g}_2$, such as ultrasound imaging [6–8]. This will be an important special case in our development, below. The fourth FOM is called G(0) and has also been related to ideal-observer detectability [12,42]. This FOM is given by

$$\begin{aligned} F_4(\mathbf{T}) &= -4 \ln \left\langle \exp \left[-\frac{1}{2} \lambda(\mathbf{v}) \right] \right\rangle_1, \\ &= -4 \ln \left\langle \exp \left[\frac{1}{2} \lambda(\mathbf{v}) \right] \right\rangle_2, \\ &= 4D_B(pr_1, pr_2). \end{aligned} \quad (14)$$

In this expression $D_B(pr_1, pr_2)$ is the Bhattacharyya distance between the two PDFs, which also has applications in information theory. The fifth and final FOM we will consider for optimizing the channel matrix is the AUC for the channelized ideal observer. It is not obvious, but it can be shown [12] that this FOM can be written as

$$F_5(\mathbf{T}) = 1 - \frac{1}{4\pi} \int_{-\infty}^{\infty} \left| \left\langle \exp \left[\left(\frac{1}{2} + i\alpha \right) \lambda(\mathbf{v}) \right] \right\rangle_2 \right|^2 \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (15)$$

This is not the usual expression for the ideal-observer AUC, but it will be useful for our calculations. The optimization problems we are considering may now be stated.

Optimization Problem j

With the normality assumption above, and for a given value of L , what channel matrix $\mathbf{T}$ will maximize $F_j(\mathbf{T})$?

A proper FOM for channel optimization should satisfy

$$F(\mathbf{MT}) = F(\mathbf{T}) \quad (16)$$

for any invertible $L \times L$ matrix $\mathbf{M}$. This is because the channels for $\mathbf{MT}$ are all linear combinations of the channels for $\mathbf{T}$, and vice-versa. We will see below that the FOMs $F_j(\mathbf{T})$ defined above all satisfy this requirement. This condition implies that the FOM can be defined as a function on the Grassmanian manifold $\mathbf{Gr}(L, M)$ [43]. Since this is a compact manifold and all the FOMs defined above are continuous functions, we know that there is a maximum for each $F_j(\mathbf{T})$. Methods have been developed for optimization on Grassmannians that use coordinate systems on the manifold itself [44]. However, the calculations described below do not follow this approach, but instead simply compute the gradient of $F_j(\mathbf{T})$ with respect to $\mathbf{T}$ and set it equal to zero. In a sense, we may regard the entries of the matrix $\mathbf{T}$ as a set of homogeneous coordinates for the corresponding point on $\mathbf{Gr}(L, M)$ and we have found that, for the functions that we are optimizing, these coordinates are useful.

In all that follows we will define the signal in the mean as $\mathbf{s} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_2$. If $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K}$, then the ideal linear observer is equivalent to one that computes the test statistic

$$\lambda_H(\mathbf{g}) = (\mathbf{K}^{-1}\mathbf{s})^\dagger \mathbf{g}, \quad (17)$$

and compares to a threshold. In the perception literature this is called the Hotelling observer and in other applications is called a prewhitened matched filter or the Fisher discriminant. This test statistic is a sufficient statistic for the detection task when the data are Gaussian and the covariances are equal. A single channel directly reduces the data to a scalar, i.e., $L = 1$. A set of channels will be optimal, in this case, if some linear combination of them is equal to the Hotelling template $(\mathbf{K}^{-1}\mathbf{s})^\dagger$. In other words, if there is an L -dimensional vector $\mathbf{c}$ such that $\mathbf{c}^\dagger \mathbf{T} = (\mathbf{K}^{-1}\mathbf{s})^\dagger$. Note that if $\mathbf{T}$ is replaced by $\mathbf{MT}$ as above, then we can replace $\mathbf{c}$ with $(\mathbf{M}^\dagger)^{-1}\mathbf{c}$ and this condition will still be satisfied. Thus, optimality is a property of a point on $\mathbf{Gr}(L, M)$ and does not depend on the specific matrix $\mathbf{T}$ used to represent that point. Having disposed of this special case, we will now assume for the remainder of this paper that $\mathbf{K}_1 \neq \mathbf{K}_2$ so that the ideal observer is a quadratic observer.

4. KL1 AND KL2

First, we compute the FOM, a computation presented in the Appendices. To describe the end result we introduce notation for the covariance matrices of the channelized data for each data class

$$\mathbf{C}_i(\mathbf{T}) = \mathbf{T} \mathbf{K}_i \mathbf{T}^\dagger. \quad (18)$$

Then we define the matrix function

$$\mathbf{R}_{12}(\mathbf{T}) = \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{C}_1(\mathbf{T}), \quad (19)$$

where the inverse exists since $\mathbf{T}$ is full rank. Next, we define the scalar function

$$H_2(\mathbf{T}) = \mathbf{s}^\dagger \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{s}. \quad (20)$$

The KL1 FOM can now be written as

$$2F_1(\mathbf{T}) = -L + \text{tr}[\mathbf{R}_{12}(\mathbf{T})] + H_2(\mathbf{T}) - \ln \det[\mathbf{R}_{12}(\mathbf{T})]. \quad (21)$$

The second term in this KL divergence is dominated by the largest eigenvalues of $\mathbf{R}_{12}(\mathbf{T})$. The third term is a Hotelling trace term in the channelized data space for the second-hypothesis covariance. The last term is dominated by the smallest eigenvalues of $\mathbf{R}_{12}(\mathbf{T})$ if we assume that they are near zero. Now, we wish to maximize this expectation over all possible channel matrices of a given dimension. To do this we will compute the gradient $\nabla F_1(\mathbf{T})$ of $F_1(\mathbf{T})$ and set it equal to zero. This gradient is an $L \times M$ matrix-valued function of $\mathbf{T}$ and is defined by the relation

$$\left. \frac{d}{dt} F_1(\mathbf{T} + t\mathbf{E}) \right|_{t=0} = \text{tr}[\mathbf{E}^\dagger \nabla F_1(\mathbf{T})], \quad (22)$$

which holds for all $L \times M$ matrices $\mathbf{E}$. We will use the notation $\Delta_1(\mathbf{T}) = \nabla F_1(\mathbf{T})$. From Appendix A we then have

$$\Delta_1(\mathbf{T}) = \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} (\mathbf{K}_1 + \mathbf{s}\mathbf{s}^\dagger) \left[\mathbf{I} - \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 \right] - \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 + \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1. \quad (23)$$

At present we do not have an analytic solution to $\Delta_1(\mathbf{T}) = \mathbf{0}$ unless $\mathbf{s} = \mathbf{0}$, which is a special case that we will examine below. The simplest iterative algorithm to try is

$$\mathbf{T}^{(n+1)} = \mathbf{T}^{(n)} + \epsilon \Delta_1(\mathbf{T}^{(n)}), \quad (24)$$

where the iterative process ends at $n = \tilde{n}$ for some stopping criteria $\mathbf{T}^{(\tilde{n})} \approx \mathbf{T}^{(\tilde{n}+1)}$. The decision variable for this CQO is then calculated by

$$\lambda_{CQO}(\mathbf{T}^{(\tilde{n})}, \mathbf{g}). \quad (25)$$

We have successfully used this type of algorithm for the J-FOM; see the Simulations section for these results. Two interesting properties of the gradient, which follow from the invariance property of the FOM, are $\Delta_1(\mathbf{T})\mathbf{T}^\dagger = \mathbf{0}$ and $\Delta_1(\mathbf{MT}) = (\mathbf{M}^\dagger)^{-1} \Delta_1(\mathbf{T})$. Note that the second property implies that the iteration depends explicitly on the starting matrix $\mathbf{T}^{(0)}$. If a different matrix $\mathbf{MT}^{(0)}$ that represents the same subspace is used, then the iteration follows a different path on the Grassmannian $\mathbf{Gr}(L, M)$. This has not proved to be a problem in implementation so far, but an algorithm that does not have this property is

$$\mathbf{T}^{(n+1)} = \mathbf{T}^{(n)} \left[\mathbf{I} + \epsilon (\mathbf{T}^{(n)})^\dagger \Delta_1(\mathbf{T}^{(n)}) \right]. \quad (26)$$

This algorithm follows the same path in $\mathbf{Gr}(L, M)$ no matter which channel matrix is used to represent the initial subspace.

When the signal is not in the mean $\mathbf{s} = \mathbf{0}$, an analytic solution to the gradient equation $\nabla F_1(\mathbf{T}) = \mathbf{0}$ is available.

Proposition 1

With the normality assumptions used in this paper, and when there is no signal in the mean, an optimal channel matrix for the KL1 FOM $F_1(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . These eigenvectors are chosen to have the L largest values of $\kappa_l - \ln \kappa_l$. If this channel matrix is denoted $\mathbf{T}_{\text{eig}}$ then the decision variable for this Eigen-CQO is calculated by

$$\lambda_{\text{CQO}}(\mathbf{T}_{\text{eig}}, \mathbf{g}). \quad (27)$$

Proof

With no signal in the mean the gradient equation can now be written as

$$[\mathbf{R}_{12}(\mathbf{T}) + \mathbf{I}] (\mathbf{T}\mathbf{K}_2\mathbf{T}^\dagger)^{-1} \mathbf{T}\mathbf{K}_2 = [\mathbf{R}_{12}(\mathbf{T}) + \mathbf{I}] (\mathbf{T}\mathbf{K}_1\mathbf{T}^\dagger)^{-1} \mathbf{T}\mathbf{K}_1. \quad (28)$$

The eigenvalues of $\mathbf{R}_{12}(\mathbf{T})$ are the same as the eigenvalues of

$$\mathbf{C}_2^{\frac{1}{2}}(\mathbf{T}) \mathbf{R}_{12}(\mathbf{T}) \mathbf{C}_2^{-\frac{1}{2}}(\mathbf{T}) = \mathbf{C}_2^{-\frac{1}{2}}(\mathbf{T}) \mathbf{C}_1(\mathbf{T}) \mathbf{C}_2^{-\frac{1}{2}}(\mathbf{T}), \quad (29)$$

which is symmetric and positive definite. These eigenvalues are therefore real and positive. This implies that $\mathbf{R}_{12}(\mathbf{T}) + \mathbf{I}$ is invertible and may be removed from the gradient equation. After doing this, taking the adjoint, and performing some additional matrix algebra we have

$$\mathbf{K}_2^{-1}\mathbf{K}_1\mathbf{T}^\dagger = \mathbf{T}^\dagger \mathbf{R}_{12}(\mathbf{T}). \quad (30)$$

Now let $\mathbf{M}$ be an invertible $L \times L$ matrix. If the channel matrix $\mathbf{T}$ produces the maximum value for $F_1(\mathbf{T})$, then the matrix $\mathbf{MT}$ does also. Setting the gradient equal to zero for this matrix gives

$$\mathbf{K}_2^{-1}\mathbf{K}_1\mathbf{T}^\dagger\mathbf{M}^\dagger = \mathbf{T}^\dagger\mathbf{M}^\dagger \left[(\mathbf{M}^\dagger)^{-1} \mathbf{R}_{12}(\mathbf{T}) \mathbf{M}^\dagger \right], \quad (31)$$

which is indeed equivalent to the original gradient equation for $\mathbf{T}$. Since, as noted above, $\mathbf{R}_{12}(\mathbf{T})$ is related to a symmetric positive definite matrix by a similarity transformation, it is diagonalizable and, as we said before, the eigenvalues are all positive real numbers. If we choose $\mathbf{M}$ to diagonalize $\mathbf{R}_{12}(\mathbf{T})$, then we have

$$\mathbf{K}_2^{-1}\mathbf{K}_1\mathbf{T}^\dagger\mathbf{M}^\dagger = \mathbf{T}^\dagger\mathbf{M}^\dagger \Lambda(\mathbf{T}), \quad (32)$$

where $\Lambda(\mathbf{T})$ is a diagonal matrix. Therefore, the columns of $(\mathbf{MT})^\dagger$ are eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ and the numbers on the main diagonal of $\Lambda(\mathbf{T})$ are eigenvalues of $\mathbf{K}_2^{-1}\mathbf{K}_1$. Now, we will replace $\mathbf{T}$ with $\mathbf{MT}$ in $F_1(\mathbf{T})$ and find that, when the gradient is zero at $\mathbf{T}$, we have

$$F(\mathbf{T}) = F(\mathbf{MT}) = \sum_l (\kappa_l - \ln \kappa_l), \quad (33)$$

where the sum is over the eigenvalues that appear in $\Lambda(\mathbf{T})$. Therefore, to maximize $F_1(\mathbf{T})$ we should choose the rows of $\mathbf{T}$ to be the L eigenvectors $\mathbf{t}_l$ of $\mathbf{K}_2^{-1}\mathbf{K}_1$ that have the L largest values for $\kappa_l - \ln \kappa_l$.

We will find similar answers for other FOMs, the only difference being which eigenvectors to choose. Note that the analytic solution depends on inverting $\mathbf{K}_2$ and finding the eigenvectors for $\mathbf{K}_2^{-1}\mathbf{K}_1$. Since these are both $M \times M$ matrices, this may be computationally difficult. On the other hand, the iterative algorithms only require inverting $\mathbf{C}_1(\mathbf{T})$ and $\mathbf{C}_2(\mathbf{T})$, which are $L \times L$ matrices. Thus, even when the signal in the mean is zero, we may want to use an iterative algorithm for practical reasons. There is a problem with local maxima for the iterative algorithms presented above and this will be discussed in detail for the J FOM, below.

With obvious notational changes, the results for the KL2 FOM can be written down from the KL1 FOM by interchanging $1 \leftrightarrow 2$ in the subscripts. We then have

$$2F_2(\mathbf{T}) = -L + \text{tr}[\mathbf{R}_{21}(\mathbf{T})] + H_1(\mathbf{T}) - \ln \det[\mathbf{R}_{21}(\mathbf{T})]. \quad (34)$$

The gradient of $F_2(\mathbf{T})$ is given by

$$\Delta_2(\mathbf{T}) = \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} (\mathbf{K}_2 + \mathbf{s}\mathbf{s}^\dagger) [\mathbf{I} - \mathbf{T}^\dagger \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1] - \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1 + \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2. \quad (35)$$

We can use this in iterative algorithms, as we did for the KL1 FOM. For no signal in the mean, there is again an analytic solution to the optimization problem.

Proposition 2

With the normality assumptions used in this paper, and when there is no signal in the mean, an optimal channel matrix for the KL2 FOM $F_2(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . These eigenvectors are chosen to have the L largest values of $\kappa_l^{-1} + \ln \kappa_l$.

Proof

We only need to note that $\mathbf{K}_1^{-1}\mathbf{K}_2 = (\mathbf{K}_2^{-1}\mathbf{K}_1)^{-1}$ and repeat the proof of Proposition 1 with the index change $1 \leftrightarrow 2$.

Note that the optimal KL1 and KL2 channel matrices both draw from the eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ but may choose different channels due to the different emphasis on using the large as opposed to the small eigenvalues of this matrix.

5. J

Since $F_3(\mathbf{T}) = F_1(\mathbf{T}) + F_2(\mathbf{T})$ the symmetrized KL divergence is given by

$$2F_3(\mathbf{T}) = -2L + \text{tr}[\mathbf{R}_{12}(\mathbf{T})] + H_2(\mathbf{T}) + \text{tr}[\mathbf{R}_{21}(\mathbf{T})] + H_1(\mathbf{T}). \quad (36)$$

The gradient of this FOM is

$$\Delta_3(\mathbf{T}) = \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} (\mathbf{K}_2 + \mathbf{s}\mathbf{s}^\dagger) \left[\mathbf{I} - \mathbf{T}^\dagger \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1 \right] + \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} (\mathbf{K}_1 + \mathbf{s}\mathbf{s}^\dagger) \left[\mathbf{I} - \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 \right]. \quad (37)$$

This gradient can be put into either of the iterative algorithms and is the one used in the simulations, for reasons to be discussed below. When there is no signal in the mean the equation $\Delta_3(\mathbf{T}) = \mathbf{0}$ can again be solved analytically.

Proposition 3

With the normality assumptions used in this paper, and when there is no signal in the mean, an optimal channel matrix for the J FOM $F_3(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . These eigenvectors are chosen to have the L largest values of $\kappa_l + \kappa_l^{-1}$.

Proof

With no signal in the mean the equation $\Delta_3(\mathbf{T}) = \mathbf{0}$ can be written as

$$[\mathbf{R}_{12}(\mathbf{T}) - \mathbf{R}_{21}(\mathbf{T})] (\mathbf{T} \mathbf{K}_1 \mathbf{T}^\dagger)^{-1} \mathbf{T} \mathbf{K}_1 = [\mathbf{R}_{12}(\mathbf{T}) - \mathbf{R}_{21}(\mathbf{T})] (\mathbf{T} \mathbf{K}_2 \mathbf{T}^\dagger)^{-1} \mathbf{T} \mathbf{K}_2. \quad (38)$$

We will assume that the matrix in the square brackets is invertible and check this assumption later. This leads to the familiar equation

$$\mathbf{K}_2^{-1} \mathbf{K}_1 \mathbf{T}^\dagger = \mathbf{T}^\dagger \mathbf{R}_{12}(\mathbf{T}). \quad (39)$$

As in Proposition 1, this leads to the result that an optimal channel matrix may be constructed from L eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ with corresponding eigenvalues κ_l . The value of the J FOM at this matrix is

$$F_3(\mathbf{T}) = \frac{1}{2} \sum_l \left(\kappa_l + \frac{1}{\kappa_l} \right) - L. \quad (40)$$

This expression leads to the last statement in the proposition.

Now we need to go back and check the invertibility assumption for the matrix $\mathbf{R}_{12}(\mathbf{T}) - \mathbf{R}_{21}(\mathbf{T})$. For positive κ , the minimum value for $\kappa + \kappa^{-1}$ occurs at $\kappa = 1$. If the corresponding eigenvector ends up in the channel matrix $\mathbf{T}$, then $\mathbf{R}_{12}(\mathbf{T}) - \mathbf{R}_{21}(\mathbf{T})$ will not be invertible, since this eigenvector would be a null vector of this matrix. However, there would be no point in including such a channel since it would not help to distinguish the two hypotheses. We can see this by writing the eigenvector equation for the channels as $\mathbf{K}_1 \mathbf{t}_l = \kappa_l \mathbf{K}_2 \mathbf{t}_l$. We can use the properties of symmetric matrices to show that these eigenvectors may be chosen to satisfy the dual orthogonality properties,

$$\mathbf{t}_l \mathbf{K}_2^\dagger \mathbf{t}_{l'} = \delta_{ll'}, \mathbf{t}_l \mathbf{K}_1^\dagger \mathbf{t}_{l'} = \kappa_l \delta_{ll'}. \quad (41)$$

Thus, the optimal channelized vector components are independent under both hypotheses, since they are Gaussian distributed in both cases. Also, if $\kappa_l = 1$, then the corresponding component has the same variance under both hypotheses. Thus, that component will not appear in the log-likelihood, and we do not need to use it. The end result is that we can remove this row from the channel matrix $\mathbf{T}$ without any loss of information and reduce the number of channels by one. If, in the process of choosing eigenvectors that maximize $\kappa_l + \kappa_l^{-1}$, we are forced to choose one with $\kappa_l = 1$, then we do not need L channels at all. In this case there would be no loss of information in using $L - 1$ channels. Of course, there is also no harm in including such a channel and, in this sense, the statement of the proposition still holds.

Note that the set of eigenvectors we choose from for optimizing J , is the same as for optimizing KL1 and KL2 but, again, we may make different choices for the L eigenvectors to use for the channel matrix $\mathbf{T}$. We will see that the last two FOMs lead to the same choice of the L eigenvectors for the optimal channel matrix as the J FOM.

6. G(0)

The expectation that appears in the G(0) FOM can be computed after some effort. We know that $F_4(\mathbf{T}) = -4 \ln I(\mathbf{T})$, where

$$I(\mathbf{T}) = \int_{\mathbb{R}^L} [p_{r1}(\mathbf{v}) p_{r2}(\mathbf{v})]^{\frac{1}{2}} d^L v. \quad (42)$$

We need to complete the square on the exponent that appears in the integrand. First, we define a symmetric matrix $\mathbf{C}(\mathbf{T})$ by $2\mathbf{C}^{-1}(\mathbf{T}) = \mathbf{C}_1^{-1}(\mathbf{T}) + \mathbf{C}_2^{-1}(\mathbf{T})$ and vector $\mathbf{c}(\mathbf{T})$ by $2\mathbf{c}(\mathbf{T}) = \mathbf{C}(\mathbf{T}) \left[\mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \bar{\mathbf{g}}_1 + \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \bar{\mathbf{g}}_2 \right]$. Next, we define the function $H(\mathbf{T})$ by

$$4H(\mathbf{T}) = 2\mathbf{c}^\dagger(\mathbf{T}) \mathbf{C}^{-1}(\mathbf{T}) \mathbf{c}(\mathbf{T}) - \bar{\mathbf{g}}_1^\dagger \mathbf{T}^\dagger \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_2^\dagger \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \bar{\mathbf{g}}_2. \quad (43)$$

Then we have the following relation,

$$[\mathbf{v} - \mathbf{c}(\mathbf{T})]^\dagger \mathbf{C}^{-1}(\mathbf{T}) [\mathbf{v} - \mathbf{c}(\mathbf{T})] - H(\mathbf{T}) = \frac{1}{2} \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_1 \right)^\dagger \mathbf{C}_1^{-1}(\mathbf{T}) \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_1 \right) + \frac{1}{2} \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_2 \right)^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_2 \right). \quad (44)$$

After performing the integral we have an expression for the G(0) FOM:

$$F_4(\mathbf{T}) = -2 \ln \det \mathbf{C}(\mathbf{T}) + \ln \det \mathbf{C}_1(\mathbf{T}) + \ln \det \mathbf{C}_2(\mathbf{T}) + H(\mathbf{T}). \quad (45)$$

Taking the gradient of this expression is very complicated and we will not pursue it further. However, when there is no signal in the mean $H(\mathbf{T}) = \mathbf{0}$ and the gradient calculation simplifies considerably. This equal mean case was considered in [29] and includes a proof that the optimal channel matrix for the G(0) FOM $F_4(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . These eigenvectors are chosen to have the L largest values of $\kappa_l + \kappa_l^{-1}$.

This means that a channel matrix that is optimal for the G(0) FOM is also optimal for the J FOM, and vice versa. This happens even though these two SFOMs are clearly different. Both of these FOMs have been put forward as good approximations to the square of the ideal-observer detectability $d_A(\mathbf{T})$ [12], which is defined by the relation

$$F_5(\mathbf{T}) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{1}{2} d_A(\mathbf{T}) \right], \quad (46)$$

where $\operatorname{erf}[\cdot]$ is the error function. It remains to be seen which of these two FOMs gives a better approximation to this detectability.

7. AUC

For the AUC FOM we will only consider the case where there is no signal in the mean. The result in this case is familiar by now.

Proposition 4

With the normality assumptions used in this paper, and when there is no signal in the mean, an optimal channel matrix for the AUC FOM $F_5(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . These eigenvectors are chosen to have the L largest values of $\kappa_l + \kappa_l^{-1}$.

Proof

We will assume that $\bar{\mathbf{g}}_1 = \bar{\mathbf{g}}_2 = \mathbf{0}$. This involves no loss of generality since we could subtract the known mean channelized data vector from the channelized data vector as a preprocessing step that would not affect the channelized ideal observer (CIO) AUC. As noted above, the CIO AUC can be written as

$$F_5(\mathbf{T}) = 1 - \frac{1}{4\pi} \int_{-\infty}^{\infty} \left| \left\langle \exp \left[\left(\frac{1}{2} + i\alpha \right) \lambda(\mathbf{v}) \right] \right\rangle_2 \right|^2 \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (47)$$

The expectation factors into two terms, one from the Gaussian normalization factors and one from the expectation integral

$$\left\langle \Lambda^{\frac{1}{2} + i\alpha}(\mathbf{v}) \right\rangle_2 = N(\alpha, \mathbf{T}) I(\alpha, \mathbf{T}). \quad (48)$$

The normalization component is

$$N(\alpha, \mathbf{T}) = \left[\frac{1}{\sqrt{(2\pi)^L \det \mathbf{C}_1(\mathbf{T})}} \right]^{\frac{1}{2} + i\alpha} \times \left[\frac{1}{\sqrt{(2\pi)^L \det \mathbf{C}_2(\mathbf{T})}} \right]^{\frac{1}{2} - i\alpha}. \quad (49)$$

The square magnitude of this factor

$$|N(\alpha, \mathbf{T})|^2 = \frac{1}{(2\pi)^L} \left[\frac{1}{\sqrt{\det \mathbf{C}_1(\mathbf{T})}} \right] \left[\frac{1}{\sqrt{\det \mathbf{C}_2(\mathbf{T})}} \right] \quad (50)$$

does not depend on α and therefore factors out of the α integral. The integral factor $I(\alpha, \mathbf{T})$ is given by

$$\int_{\mathbb{R}^L} \exp \left\{ -\frac{1}{4} \mathbf{v}^\dagger [\mathbf{C}_1^{-1}(\mathbf{T}) + \mathbf{C}_2^{-1}(\mathbf{T})] \mathbf{v} - \frac{i\alpha}{2} \mathbf{v}^\dagger [\mathbf{C}_1^{-1}(\mathbf{T}) - \mathbf{C}_2^{-1}(\mathbf{T})] \mathbf{v} \right\} d\mathbf{v}. \quad (51)$$

Now we replace $\mathbf{T}$ with $\mathbf{MT}$. This does not change the CIO AUC since the variable $\mathbf{v}$ in the integral will be replaced by $\mathbf{u} = \mathbf{M}\mathbf{v}$ and the Jacobian factor from the change of variable in $I(\alpha, \mathbf{T})$ will cancel out with a $|\det(\mathbf{M})|$ factor from $N(\alpha, \mathbf{T})$. We may choose $\mathbf{M}$ to diagonalize both channel covariance matrices with $\mathbf{MC}_2(\mathbf{T})\mathbf{M}^\dagger = \mathbf{I}$ and $\mathbf{MC}_1(\mathbf{T})\mathbf{M}^\dagger = \mathbf{D}(\mathbf{T})$. The diagonal matrix $\mathbf{D}(\mathbf{T})$ has the eigenvalues of $\mathbf{R}_{12}(\mathbf{T})$ along the diagonal. These diagonal entries are also the eigenvalues of $\mathbf{C}_2^{-\frac{1}{2}}(\mathbf{T}) \mathbf{C}_1(\mathbf{T}) \mathbf{C}_2^{-\frac{1}{2}}(\mathbf{T})$, which is symmetric and positive definite. Therefore, the diagonal entries of $\mathbf{D}(\mathbf{T})$ are real and positive. We will call these numbers μ_l for $l = 1, \dots, L$. Now we have for the normalization component

$$|N(\alpha, \mathbf{MT})|^2 = \frac{1}{(2\pi)^L \sqrt{\prod_{l=1}^L \mu_l}}. \quad (52)$$

We may also write this equation as

$$|N(\alpha, \mathbf{MT})|^2 = \frac{1}{(2\pi)^L \sqrt{\det \mathbf{R}_{12}(\mathbf{T})}}. \quad (53)$$

For the integral factor we have

$$I(\alpha, \mathbf{MT}) = \int_{\mathbb{R}^L} \exp \left[-\frac{1}{4} \sum_{l=1}^L \left(\frac{1}{\mu_l} + 1 \right) u_l^2 - \frac{i\alpha}{2} \sum_{l=1}^L \left(\frac{1}{\mu_l} - 1 \right) u_l^2 \right] d^L \mathbf{u}. \quad (54)$$

This is a product of one dimensional integrals and can be computed analytically with the result

$$I(\alpha, \mathbf{MT}) = \pi^{\frac{L}{2}} \prod_{l=1}^L \left[\frac{1}{4} \left(\frac{1}{\mu_l} + 1 \right) + \frac{i\alpha}{2} \left(\frac{1}{\mu_l} - 1 \right) \right]^{-\frac{1}{2}}. \quad (55)$$

We may also write this equation as

$$I(\alpha, \mathbf{MT}) = \frac{(2\pi)^{\frac{L}{2}}}{\sqrt{\det \left[\left(\frac{1}{2} + i\alpha \right) \mathbf{R}_{21}(\mathbf{T}) + \left(\frac{1}{2} - i\alpha \right) \mathbf{I} \right]}}. \quad (56)$$

since $\mathbf{R}_{21}(\mathbf{T}) = \mathbf{R}_{12}^{-1}(\mathbf{T})$. Combining the results so far we have

$$\left| \left\langle \Lambda^{\frac{1}{2} + i\alpha}(\mathbf{v}) \right\rangle_2 \right|^2 = [P(\alpha, \mathbf{T})]^{-\frac{1}{2}} \quad (57)$$

with $P(\alpha, \mathbf{T}) = \det \mathbf{P}(\alpha, \mathbf{T})$ and $\mathbf{P}(\alpha, \mathbf{T})$ is the matrix given by

$$\left[\left(\frac{1}{2} + i\alpha \right) \mathbf{R}_{12}(\mathbf{T}) + \left(\frac{1}{2} - i\alpha \right) \mathbf{I} \right] \times \left[\left(\frac{1}{2} + i\alpha \right) \mathbf{R}_{21}(\mathbf{T}) + \left(\frac{1}{2} - i\alpha \right) \mathbf{I} \right]. \quad (58)$$

This matrix simplifies to

$$\mathbf{P}(\alpha, \mathbf{T}) = \left(\frac{1}{4} + \alpha^2 \right) [\mathbf{R}_{12}(\mathbf{T}) + \mathbf{R}_{21}(\mathbf{T})] + 2 \left(\frac{1}{4} - \alpha^2 \right) \mathbf{I}. \quad (59)$$

We now have

$$F_5(\mathbf{T}) = 1 - \frac{1}{4\pi} \int_{-\infty}^{\infty} [P(\alpha, \mathbf{T})]^{-\frac{1}{2}} \frac{d\alpha}{\alpha^2 + \frac{1}{4}}, \quad (60)$$

which we can also write as

$$F_5(\mathbf{T}) = 1 - \frac{1}{4\pi} \int_{-\infty}^{\infty} \prod_{l=1}^L \left[\left(\frac{1}{4} + \alpha^2 \right) \left(\mu_l + \frac{1}{\mu_l} \right) + 2 \left(\frac{1}{4} - \alpha^2 \right) \right]^{-\frac{1}{2}} \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (61)$$

To compute the gradient of $F_5(\mathbf{T})$ we need the gradient of $P(\alpha, \mathbf{T})$. This can be found in the Appendices. The end result of setting $\mathbf{s}(\mathbf{T}) = \mathbf{0}$ is the familiar equation

$$\mathbf{K}_2^{-1} \mathbf{K}_1 \mathbf{T}^\dagger = \mathbf{T}^\dagger \mathbf{R}_{12}(\mathbf{T}). \quad (62)$$

We may conclude, as before, that the μ_l are eigenvalues of $\mathbf{K}_2^{-1} \mathbf{K}_1$ and that we may take the rows of the channel matrix $\mathbf{T}$ to be the corresponding eigenvectors.

We will replace μ_l with κ_l now and pick the κ_l to minimize the integral expression

$$\int_{-\infty}^{\infty} \prod_{l=1}^L \left[\left(\frac{1}{4} + \alpha^2 \right) \left(\kappa_l + \frac{1}{\kappa_l} \right) + 2 \left(\frac{1}{4} - \alpha^2 \right) \right]^{-\frac{1}{2}} \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (63)$$

The corresponding eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ will give the rows of the optimum channel matrix. This integral expression can also be written as

$$\int_{-\infty}^{\infty} \prod_{l=1}^L \left[\frac{1}{4} \frac{(\kappa_l + 1)^2}{\kappa_l} + \alpha^2 \frac{(\kappa_l - 1)^2}{\kappa_l} \right]^{-\frac{1}{2}} \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (64)$$

Now, define

$$y_l = \frac{4\kappa_l}{(\kappa_l + 1)^2}, \quad (65)$$

and write the integral expression we want to minimize by our choice of the κ_l as

$$\int_{-\infty}^{\infty} \prod_{l=1}^L \sqrt{\frac{y_l}{1 + 4\alpha^2(1 - y_l)}} \frac{d\alpha}{\alpha^2 + \frac{1}{4}}. \quad (66)$$

Now, there are two facts we can use to solve our problem. A little algebra shows that $0 < y_l < 1$ since all of the κ_l are positive. Within this range the function

$$f(y) = \frac{y}{1 + 4\alpha^2(1 - y)} \quad (67)$$

is monotonic, increasing for any α . Therefore, we should choose the eigenvalues κ_l that produce the L minimum values for y_l . But

$$\frac{1}{y_l} = \frac{1}{4} \left(\kappa_l + \frac{1}{\kappa_l} + 2 \right). \quad (68)$$

Therefore, we may conclude that to maximize $F_5(\mathbf{T})$ we should choose the rows of $\mathbf{T}$ to be the eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ whose corresponding eigenvalues κ_l produce the L largest values for $\kappa_l + \kappa_l^{-1}$. If we go back to Appendix B and examine the matrices $\mathbf{U}_1(\mathbf{T})$ and $\mathbf{U}_2(\mathbf{T})$ for this channel matrix we find that they are diagonal and invertible as long as $\kappa_l \neq 1$. If any of these eigenvalues are unity, then this situation can be dealt with as in Proposition 3.

The optimal channel matrix for the AUC FOM is the same optimal channel matrix we arrived at for the J and G(0) FOMs. Thus, these quantities can be regarded as surrogate FOMs for the CIO AUC in the sense that they are optimized at the same channel matrix that optimizes the AUC FOM, which is the preferred FOM since it is directly related to task performance. On the other hand the J and G(0) FOMs are also easier to compute than the AUC FOM. The J FOM, in particular, is much easier to deal with than the AUC FOM and, since it leads to the same optimal channel matrix, we have used the J FOM in our simulations. This leads to an examination of the possibility of local maxima for the J FOM, which we examine in the next section.

8. LOCAL MAXIMA FOR J WITH NO SIGNAL IN THE MEAN

To determine the local maxima we need the Hessian for $F_3(\mathbf{T})$. This Hessian evaluated at $\mathbf{T}$ is a linear operator $\mathcal{D}(\mathbf{T})$ from the space of $L \times M$ matrices to itself. If $\mathbf{E}$ is an arbitrary $L \times M$ matrix, then $\mathcal{D}(\mathbf{T})$ is determined uniquely as the operator that satisfies

$$\left. \frac{d^2}{dt^2} F_3(\mathbf{T} + t\mathbf{E}) \right|_{t=0} = \text{tr} \left[\mathbf{E}^\dagger \mathcal{D}(\mathbf{T}) \mathbf{E} \right] \quad (69)$$

for all $\mathbf{E}$. In this notation $\mathcal{D}(\mathbf{T}) \mathbf{E}$ is the $L \times M$ matrix that results from the linear operator $\mathcal{D}(\mathbf{T})$ acting on $\mathbf{E}$. Specifically, we are interested in the eigenvalues of the Hessian at the points where the gradient of $F_3(\mathbf{T})$ is zero.

Proposition 5

With the normality assumptions used in this paper, and when there is no signal in the mean, any local maximum for the J FOM $F_5(\mathbf{T})$ can be constructed by using L eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ for the channels, with corresponding eigenvalues κ_l . For an integer N with $0 < N < L$ these eigenvectors must be chosen to have the N largest values of κ_l that are also > 1 , and the $L - N$ smallest values of κ_l that are also < 1 . Thus, barring degeneracies, there are no more than $L + 1$ local maxima, one of which is the global maximum.

Proof

We will fix the channel matrix $\mathbf{T}$ and the perturbation matrix $\mathbf{E}$ and write $\mathbf{C}_i(\mathbf{T} + t\mathbf{E}) = \mathbf{C}_i(t)$ and $\mathbf{R}_{ij}(\mathbf{T} + t\mathbf{E}) = \mathbf{R}_{ij}(t)$. Then we have $F_3(t) = \text{tr}[\mathbf{R}_{12}(t)] + \text{tr}[\mathbf{R}_{21}(t)]$ with the relations $\mathbf{C}_2(t)\mathbf{R}_{12}(t) = \mathbf{C}_1(t)$ and $\mathbf{C}_1(t)\mathbf{R}_{21}(t) = \mathbf{C}_2(t)$. The basic ingredients of the calculation are the following equations:

$$\mathbf{C}_j(0) = \mathbf{T} \mathbf{K}_j \mathbf{T}^\dagger, \quad (70)$$

$$\mathbf{C}'_j(0) = \mathbf{E} \mathbf{K}_j \mathbf{T}^\dagger + \mathbf{T} \mathbf{K}_j \mathbf{E}^\dagger, \quad (71)$$

$$\mathbf{C}''_j(0) = 2\mathbf{E} \mathbf{K}_j \mathbf{E}^\dagger, \quad (72)$$

$$\mathbf{R}'_{ij}(0) = \mathbf{C}_j^{-1}(0) \left[\mathbf{C}'_i(0) - \mathbf{C}'_j(0) \mathbf{R}_{ij}(0) \right], \quad (73)$$

$$\mathbf{R}''_{ij}(0) = \mathbf{C}_j^{-1}(0) \left[\mathbf{C}''_i(0) - \mathbf{C}''_j(0) \mathbf{R}_{ij}(0) - 2\mathbf{C}'_j(0) \mathbf{R}'_{ij}(0) \right], \quad (74)$$

$$F''_3(0) = \text{tr} \left[\mathbf{R}''_{12}(0) \right] + \text{tr} \left[\mathbf{R}''_{21}(0) \right]. \quad (75)$$

With these equations we can compute $\text{tr} \left[\mathbf{E}^\dagger \mathcal{D}(\mathbf{T}) \mathbf{E} \right]$ for any channel matrix $\mathbf{T}$ and perturbation matrix $\mathbf{E}$.

When we are at a point where the gradient is zero, then, as noted above, we may assume that $\mathbf{K}_1 \mathbf{T}^\dagger = \mathbf{K}_2 \mathbf{T}^\dagger \Lambda$ where Λ is a diagonal matrix. We also may assume the dual orthogonality conditions [see Eq. (41)] on the rows of $\mathbf{T}$, which we can write as $\mathbf{C}_1(0) = \Lambda$ and $\mathbf{C}_2(0) = \mathbf{I}$.

To proceed further it is useful to define the matrix $\mathbf{T}_0$ as the $(M-L) \times M$ matrix whose rows form the set of eigenvectors of $\mathbf{K}_2^{-1} \mathbf{K}_1$ that do not appear in the given channel matrix $\mathbf{T}$. We also define Λ_0 as the diagonal matrix with the corresponding eigenvalues κ_{0m} on the diagonal. Now we have the orthogonality relations $\mathbf{T}_0 \mathbf{K}_j \mathbf{T}^\dagger = \mathbf{0}$ and $\mathbf{T} \mathbf{K}_j \mathbf{T}_0^\dagger = \mathbf{0}$. We may also assume that these eigenvectors satisfy $\mathbf{T}_0 \mathbf{K}_1 \mathbf{T}_0^\dagger = \Lambda_0$ and $\mathbf{T}_0 \mathbf{K}_2 \mathbf{T}_0^\dagger = \mathbf{I}$. We may then write a decomposition of the perturbation matrix $\mathbf{E} = \mathbf{B} \mathbf{T} + \mathbf{B}_0 \mathbf{T}_0$ where $\mathbf{B}$ is $L \times L$ and $\mathbf{B}_0$ is $L \times (M-L)$. Then $\mathbf{R}_{12}(0) = \Lambda$ and $\mathbf{R}_{21}(0) = \Lambda^{-1}$ and

$$\mathbf{C}'_1(0) = \mathbf{B} \Lambda + \Lambda \mathbf{B}^\dagger \quad (76)$$

$$\mathbf{C}'_2(0) = \mathbf{B} + \mathbf{B}^\dagger \quad (77)$$

$$\mathbf{C}''_1(0) = 2\mathbf{B} \Lambda \mathbf{B}^\dagger + 2\mathbf{B}_0 \Lambda_0 \mathbf{B}_0^\dagger \quad (78)$$

$$\mathbf{C}''_2(0) = 2\mathbf{B} \mathbf{B}^\dagger + 2\mathbf{B}_0 \mathbf{B}_0^\dagger. \quad (79)$$

Placing these in the formula above for $F''_3(0)$ we find that all of the terms with $\mathbf{B}$ in them cancel, as they should since a perturbation of the form $\mathbf{E} = \mathbf{B} \mathbf{T}$ would give $F_3(t) = F_3(0)$ for all t in an interval about 0. Therefore, we may restrict our perturbations to those of the form $\mathbf{E} = \mathbf{B}_0 \mathbf{T}_0$, since these are the directions that move us away from the point corresponding to $\mathbf{T}$ on the Grassmann manifold $\mathbf{Gr}(L, M)$. The final result is

$$\text{tr} \left[\mathbf{E}^\dagger \mathcal{D}(\mathbf{T}) \mathbf{E} \right] = 2 \text{tr} \left[\left(\Lambda^{-1} - \Lambda \right) \left(\mathbf{B}_0 \mathbf{B}_0^\dagger - \Lambda^{-1} \mathbf{B}_0 \Lambda_0 \mathbf{B}_0^\dagger \right) \right]. \quad (80)$$

Using the orthogonality relations of the eigenvectors we can also write this as

$$\text{tr} \left[\mathbf{E}^\dagger \mathcal{D}(\mathbf{T}) \mathbf{E} \right] = 2 \text{tr} \left[(\mathbf{B}_0 \mathbf{T}_0)^\dagger (\Lambda^{-1} - \Lambda) (\mathbf{B}_0 - \Lambda^{-1} \mathbf{B}_0 \Lambda_0) \mathbf{T}_0 \mathbf{K}_2 \right]. \quad (81)$$

Therefore, the Hessian operator at a point on $\mathbf{Gr}(L, M)$ corresponding to a channel matrix $\mathbf{T}$ where the gradient of $F_3(\mathbf{T})$ is zero is given by

$$\mathcal{D}(\mathbf{T}) \mathbf{B}_0 \mathbf{T}_0 = (\Lambda^{-1} - \Lambda) (\mathbf{B}_0 - \Lambda^{-1} \mathbf{B}_0 \Lambda_0) \mathbf{T}_0 \mathbf{K}_2. \quad (82)$$

We want to know the eigenvalues of this operator.

First, we will show that, without loss of generality, we may assume that $\mathbf{K}_2$ is the identity matrix. If this is not the case define a new data vector by $\mathbf{g}' = \mathbf{K}_2^{-\frac{1}{2}} \mathbf{g}$. Then, in the obvious notation, we have new covariance matrices $\mathbf{K}'_2 = \mathbf{I}$ and $\mathbf{K}'_1 = \mathbf{K}_2^{-\frac{1}{2}} \mathbf{K}_1 \mathbf{K}_2^{-\frac{1}{2}}$. If $\mathbf{T}$ is a channel matrix, define $\mathbf{T}' = \mathbf{T} \mathbf{K}_2^{\frac{1}{2}}$. Then we have for channel outputs $\mathbf{v} = \mathbf{T} \mathbf{g} = \mathbf{T}' \mathbf{g}'$. It may also be easily verified that

$$\begin{aligned} F'_3(\mathbf{T}') &= \text{tr} \left[(\mathbf{T}' \mathbf{K}'_2 \mathbf{T}'^\dagger)^{-1} (\mathbf{T}' \mathbf{K}'_1 \mathbf{T}'^\dagger) \right] + \text{tr} \left[(\mathbf{T}' \mathbf{K}'_1 \mathbf{T}'^\dagger)^{-1} (\mathbf{T}' \mathbf{K}'_2 \mathbf{T}'^\dagger) \right] \\ &= F_3(\mathbf{T}), \end{aligned} \quad (83)$$

which we would expect since the channel outputs are the same. Therefore, $\mathbf{T}$ is a local maximum for $F_3(\mathbf{T})$ if and only if $\mathbf{T}'$ is a local maximum for $F'_3(\mathbf{T}')$. Also note that the eigenvalues of $\mathbf{K}_2^{-1} \mathbf{K}_1$ are the same as those for $(\mathbf{K}'_2)^{-1} \mathbf{K}'_1$, although the eigenvectors are different.

Dropping the primes we will now assume that $\mathbf{K}_2 = \mathbf{I}$. Let us define an $L \times (M - L)$ matrix by $[\mathbf{B}_0^{pq}]_{lm} = \delta_{pl} \delta_{qm}$. We then have

$$\mathcal{D}(\mathbf{T}) \mathbf{B}_0^{pq} \mathbf{T}_0 = \left(\frac{1}{\kappa_p} - \kappa_p \right) \left(1 - \frac{\kappa_{0q}}{\kappa_p} \right) \mathbf{B}_0^{pq} \mathbf{T}_0. \quad (84)$$

We have a local maximum if and only if all of the eigenvalues of $\mathcal{D}(\mathbf{T})$ are negative. Therefore, the necessary and sufficient conditions for a local maximum are:

- 1) For all $\kappa_p < 1$ we have $\kappa_{0q} > \kappa_p$ for all q
- 2) For all $\kappa_p > 1$ we have $\kappa_{0q} < \kappa_p$ for all q .

Therefore, the κ_p at a local maximum must consist of the N largest eigenvalues of $\mathbf{K}_2^{-1} \mathbf{K}_1$ that are also > 1 , and the $L - N$ smallest eigenvalues that are also < 1 , for some N with $0 \leq N \leq L$. Thus, barring degeneracies, there are no more than $L + 1$ local maxima, one of which is the global maximum.

In principle running a gradient-based algorithm repeatedly from different starting points would eventually locate all of the local maxima, and thus be able to determine the global maximum. The corresponding channel matrix will also provide a global maximum for the $G(0)$ and AUC FOMs.

9. SIMULATIONS

In this section, the AUC of the J-CQO calculated from a gradient-based search [see Eq. (25)] is compared to three conventional observers: ideal [see Eq. (5)], Hotelling [see Eq. (17)], and Eigen-CQO [see Eq. (27)] under several simulated imaging conditions which correspond to various differences in the first- and second-order statistics of the two classes. The two classes of images are simulated on a 50×50 pixel grid. Here, $M = 50^2$ and this number is selected to accommodate the computational time required for the Ideal, Hotelling, and Eigen-CQO which all require an $M \times M$ matrix inverse. By contrast the J-CQO only requires inversion of $L \times L$ matrices where L is the number of channels.

The covariance matrix of each class is chosen to be circulant with a Gaussian kernel and parameterized by a correlation length as in

$$[\mathbf{K}_i]_{\mathbf{n},\mathbf{m}} = \left(\frac{\bar{b}}{SNR} \right)^2 \times \exp \left(- \frac{((n_x - m_x) \bmod 50)^2 + ((n_y - m_y) \bmod 50)^2}{2\sigma_i^2} \right) \quad (85)$$

where σ_i is the correlation length of the i th class, $\mathbf{n}$ and $\mathbf{m}$ are two-dimensional pixel indices, the SNR is set to 100 and the average value of the background $\bar{b}$ is 25 through the simulation study. The mean of the first class is constant across the image plane, i.e., all elements of $\bar{\mathbf{g}}_1$ equal $\bar{b}$. The correlation length of the second class is always $\sigma_2 = 0.3$ pixels. This relatively simple model of the covariance matrices is chosen to provide an analytic solution for the Eigen-CQO decision variable.

The observers' performance is compared for different choices of $\bar{\mathbf{g}}_2$ and σ_1 . The case of signal present in the mean is simulated by

$$\bar{\mathbf{g}}_2 = \bar{\mathbf{g}}_1 + A_s \mathbf{s} \quad (86)$$

where the support of $\mathbf{s}$ is one pixel in the middle of the image. The signal strength is expressed as contrast which is defined as $(A_s + b)/b$; a contrast of one indicates the signal is absent. Differences in the covariance matrices of the two classes are expressed as the difference in their correlation lengths, $\sigma = \sigma_1 - \sigma_2$. Example images from each class are displayed side by side in Fig. 1. In Fig. 1(a) $\sigma = 0.2$ pixels and in Fig. 1(b) $\sigma = 0.5$ pixels.

The AUC is estimated from the percent correct of 1,000 independent sample images [2]. Five independent trials of the AUC estimates are calculated and the mean and standard deviation are displayed as error plots in Figs. 2, 4, and 5. The k th sample image from the i th class is computed from

$$\mathbf{g}_i^k = \bar{\mathbf{g}}_i + \mathbf{K}_i^{1/2} \mathbf{w} \quad (87)$$

where $\mathbf{w}$ is a zero mean image of uncorrelated white noise of unit variance.

In Fig. 2 the AUC is calculated as a function of correlation length differences between the two classes while the number of channels, L , remains constant. This is repeated for both signal absent and present in the image means; see Figs. 2(a) and 2(b) respectively. In Fig. 2(a) all observers achieve an AUC of 0.5 as expected when $\sigma = 0$ and $\mathbf{g}_1 = \mathbf{g}_2$ since there is no difference between the two classes in this case. The AUC of the ideal observer saturates to 1.0 when a relatively small correlation length difference is introduced. The Hotelling template is zero when the means of the two classes are equal so its AUC remains at 0.5 and is invariant to the correlation length difference. As σ increases the performance of both the J-CQO and Eigen-CQO increase. The AUC of the Eigen-CQO yields better performance as predicted in Proposition 4 but in this simulation the difference is less than about 0.02 in AUC and the performance gap decreases as σ increases. Both observers saturate to an AUC of 1.0 when σ is sufficiently large.

As shown in Fig. 2(b) the Eigen-CQO is not necessarily better than the J-CQO when the image means are different. Here, the AUC is evaluated over the same range of σ but the AUC is not 0.5 at $\sigma = 0$ since $\mathbf{g}_1 \neq \mathbf{g}_2$. At $\sigma = 0$ all observers have approximately the same AUC of 0.68 except the Eigen-CQO which is close to 0.50. This is expected since the Eigen-CQO uses only second-order statistics of the image which are identical in this case. The Hotelling observer AUC is not expected to change as a function of σ since it only uses first-order statistics of the images (i.e., an affine transform) to calculate a decision variable. The change in Hotelling AUC in Fig. 2(b) is due to sample statistics from estimating the AUC using 5 independent trials of 1,000 images from each class. As expected the ideal observer's AUC is equal to or greater than the AUC for the signal absent case since this task is easier. Again, the Eigen-CQO and J-CQO performance both increase with σ but when the signal is present in the mean the J-CQO performance is better for the more difficult tasks. In addition to this performance increase a critical advantage of J-CQO over Eigen-CQO is tractability of the calculation for high-dimensional images. Here, images are simulated on only a 50×50 grid to allow reasonable calculation times of the Eigen-CQO. For larger images calculating the channels for CQO from the gradient-based method would remain feasible while the $M \times M$ matrix inverse required for the Eigen-CQO would become infeasible.

To understand the computation of the Eigen-CQO, Fig. 3 presents the eigenvalues of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for several values of σ . Recalling Eq. (27) the channels for Eigen-CQO are computed by selecting the eigenvectors that correspond to the largest values of $\kappa + 1/\kappa$. Comparing the eigenspectrum in Fig. 3 for six σ values reveals that as the difference between the two classes increases the abundance of small eigenvalues that contribute to the $\kappa + 1/\kappa$ rank-ordering also increases. If the number of channels were increased eventually larger eigenvalues would also be included in the $\kappa + 1/\kappa$ rank-ordering. These trends are specific to the circulant Gaussian covariance matrices used in this example [see Eq. (86)]. Note that according to Eq. (41) each κ_l is the ratio between the l th channel output variances under each of the two classes. If these variances are equal then the channel output offers no discrimination when $\mathbf{g}_1 = \mathbf{g}_2$. The maximum discrimination occurs when this ratio is either

large or small compared to one. In theory different combinations of eigenvalues from the high and low end of the eigenspectrum result in 26 (i.e., $L + 1$) local maxima as discussed in Section 8. In this simulation study independent random samples were used as the initial solution in the iterative algorithm (see Fig. 6 for both mean equal and unequal cases) and resulted in different values of J which confirms the presence of local maxima.

To investigate the change in observer performance as a function of signal strength, Fig. 4 shows all observer's AUC as the signal contrast changes. The number of channels for the J-CQO and the Eigen-CQO is held constant. Per our definition of signal contrast, 1.0 is equivalent to signal absent, hence the Hotelling AUC is 0.5 at this point. However at this point the J-CQO is about 0.81 since σ is held constant at 0.15 pixels. The J-CQO AUC is higher than Hotelling, both increase with signal contrast, and their values become increasingly similar as the contrast is increased. The AUC for both observers eventually saturates to 1.0 as the contrast is increased. The Eigen-CQO is constant with respect to signal contrast (within fluctuations due to sample statistics) since it uses only second-order statistics to calculate a decision variable and the second-order statistics are not changing. Another critical advantage of the J-CQO over the Eigen-CQO is that both first- and second-order statistics of the images are combined to calculate a decision variable.

Figure 5 is an investigation into the effect of number of channels on Eigen-CQO and J-CQO for both (a) signal absent and (b) present in the mean. The number of channels range from 1 to 50 which corresponds to a compression ratio of 2500 and 50, respectively. As expected, the ideal observer's AUC is saturated to 1.0 and the Hotelling AUC is 0.50 when the signal is absent in the mean. In Fig. 5(a) for the signal absent case both the Eigen-CQO and J-CQO AUC increase with number of channels, the rate of this increase is approximately the same for both, and the Eigen-CQO performs slightly better as predicted in Proposition 4. A very weak signal contrast of 1.01 is used in Fig. 5(b) for the signal present case. Here, the J-CQO's AUC is higher than the Eigen-CQO over the selected range of number of channels. As the number of channels increases the Eigen-CQO and J-CQO performance become increasingly similar. Since the J-CQO implementation requires selecting the number of channels prior to image processing the sensitivity of performance to number of channels will be an important quantity to determine experimentally when using the J-CQO in a new imaging application.

The channels of the J-CQO are calculated using a nonlinear iterative algorithm with an additive update term in Eq. (24). For both signal present and absent in the mean, Figs. 6(c) and 6(d) show the magnitude of the gradient, with respect to iteration number, multiplied by a scalar step-size factor ε which is used to scale the additive update. The calculation was repeated three times for independent random initial values for the channel matrix entries. Smooth convergence toward a negligible gradient is observed for all three runs. Figs. 6(a) and 6(b) shows the scalar-valued quantity to be maximized (i.e., J) as a function of iteration number, also for signal absent and present and three independent runs each. All three runs converge quickly over this range of iterations but to slightly different values of J for each run. For the case of signal absent in the mean [Fig. 6(a)] the three trial runs of J-CQO are compared to the Eigen-CQO value of J which, as expected, is greater.

Figure 7(a) is the eigenspectrum of $\mathbf{R}_{12}$ for the equal means case. Here, $L = 25$ and for the equal means case the 25 eigenvalues for each independent run are so nearly equal that they are plotted on top one another in the figure. As shown in Eq. (40) the value of $(1/2)(\kappa_l + (1/\kappa_l)) - L$ is equal to the contribution that the channel corresponding to the eigenvalue κ_l makes to the J FOM. Therefore, the plot of $(\kappa_l + (1/\kappa_l))$ versus κ_l can be interpreted as rank-ordering the importance of each eigenvector of $\mathbf{R}_{12}$ with respect to the J FOM. It is interesting to note from Eq. (40) that if one channel is added with a value of $(\kappa_l + (1/\kappa_l))$ equal to 2 then the value of J would not change by adding that channel. In Fig. 7(a) none of these values are equal to 2 and therefore deleting any of the channels would decrease the value of J.

Figure 7(b) is the eigenspectrum of $\mathbf{R}_{12}$ for the unequal means case. When the two classes have different mean values the eigenspectrum changes, as compared to the equal means case, most notably by a single eigenvalue that appears close to one. The disagreement between each independent run is most notable in the selection of this eigenvalue closest to one. The corresponding channel is the matrix-vector product of $\mathbf{T}^\dagger$ and the eigenvector. For one of the independent runs this channel is given in Fig. 7(c) to show the dependence of the J-CQO channels on the difference in the means between the two classes, which is isolated in a single pixel at the center of the image.

In Fig. 8 the top row displays the channels solutions (i.e., the rows of the matrix $\mathbf{T}$) for the J-CQO iterative algorithm for three values of the correlation length difference between the two classes with equal means. These solutions are nonunique since $\mathbf{T}$ and $\mathbf{MT}$ give the same value for J [see Eq. (16)]. To allow visual comparisons between different σ the bottom row of Fig. 8 shows the eigenvectors of $\mathbf{R}_{12}$ [see Eq. (19)] back projected by $\mathbf{T}^\dagger$. This corresponds to choosing the matrix $\mathbf{M}$ that diagonalizes $\mathbf{R}_{12}$ therefore the channel matrices in the top row produce the same value of J as the corresponding channel matrices in the bottom row. The Eigen-CQO channels, which are optimal for the equal means case, are expected to be spatial frequency channels since the matrix $\mathbf{K}_1^{-1}\mathbf{K}_2$ is stationary in this example [see Eq. (86)]. The appearance of the J-CQO channels in the bottom row of Fig. 8 are encouraging since they exhibit obvious frequency and orientation characteristics.

10. CONCLUSIONS

In this paper, we have shown a promising new method for computing optimized channels for quadratic observers utilizing high-dimensional image data. We have assumed Gaussian statistics for the channelized image data, but the method for calculating channels is applicable in general. Gradient-based algorithms for determining the channels have been presented for five different information-based FOMs: KL1, KL2, J, G(0), and AUC. Analytic solutions for the optimum channels for each of the five FOMs have been derived for the case of equal mean data for both classes. The optimum channels for the J, G(0), and AUC FOMs under the equal mean condition are the same. This result is critical for implementing J as a surrogate FOM for AUC since the J FOM is much easier to compute. Since the algorithm is gradient-based, local maxima are an important consideration. For the J FOM, with the equal mean condition, we have shown that there are no more than $L + 1$ local maxima where L is the number of channels. In principle all of the local maxima of the

equal-mean J FOM could be located by using multiple random starting locations in the algorithm. In a simulation study we have calculated the performance of ideal and Hotelling observers and compared these to CQO performance. Optimal channel selection for the CQO has been calculated from two different methods: eigenanalysis and an iterative gradient-based algorithm. The gradient-based algorithm is an important new contribution since it makes optimal CQO implementation possible without inverting an $M \times M$ matrix for eigenanalysis.

Implementing the CQO in a new imaging application requires a set of channels and the first- and second-order statistics of channelized image data from both classes. A common technique to estimate the channelized first- and second-order statistics is from finite training data for both classes. Here, the dimensionality reduction from M number of original measurements to L number of channels is a critical advantage of the J-CQO method since estimating first- and second-order statistics from uncompressed image data would require larger sample sizes. Another important advantage of the CQO is that we only need to invert two $L \times L$ matrices instead of inverting two $M \times M$ matrices. In this simulation study we have shown excellent CQO performance for compression ratios (M/L), on the order of 100. In practice, this compression ratio will depend upon the difficulty level of the task and performance requirements of the observer.

The J-CQO algorithm has the advantages of the CQO plus the property that the channels can be estimated using image samples from both classes. This can make binary classification possible in an exploratory imaging application where the relationship between the signal to be detected and the image data is not well understood. In such an application the J-CQO channel solution gives insight into which measurements are most useful for the detection task. The iterative solution for the J-CQO channels [see Eqs. (38) and (24)] does not require estimates of the image covariances $\mathbf{K}_1$ and $\mathbf{K}_2$ as an intermediate step. At each step in the iteration, training images from each class are channelized (i.e., compressed) using the candidate channel matrix solution. The sample means and sample covariance matrices for the channel outputs are computed. The sample cross-covariance matrices between the channel outputs and the original data are also computed (these are the matrices $\mathbf{TK}_1$ and $\mathbf{TK}_2$). The sample channelized means, sample cross-covariance matrices, and inverses of the sample channelized covariance matrices are then used to calculate the gradient of J with respect to the channel matrix $\mathbf{T}$. As long as the number of independent training images from each class exceeds the number of channels, the sample channelized covariance matrices will be full rank and therefore invertible. The novelty of this approach is that a prior description of the signal is not required. Instead, training data from each class can be used to estimate the channels.

We are currently working on analytic results for the case of nonequal means. Our simulation results showed that J-CQO performed better than Eigen-CQO for the case of nonequal means; see Figs. 2(b) and 5(b). In this situation, results in Fig. 7(c) indicate that to, achieve optimal-performance, Hotelling-type channels that are not eigenvectors of $\mathbf{K}_2^{-1}\mathbf{K}_1$ are necessary. We are also developing analytic results and algorithms for optimizing channelized observer performance for estimation tasks. In this case, we use the Fisher

information as a FOM. To meet the demands of a wider variety of imaging applications we will extend these CQO methods to classification tasks where there are more than two classes. Results in [45] indicate that human visual adaptation is nonlinear, this motivates testing the predictive power of J-CQO performance to human performance for texture discrimination.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under CHE-1313892 and the National Institutes of Health under R01-EB000803 and P41-EB002035. The authors would like to thank Harrison Barrett and Luca Caucci for their valuable feedback. The comments by our anonymous reviewers improved this manuscript and we appreciate their work.

APPENDIX A: KL1 AND KL2

To compute $F_1(\mathbf{T})$, first we find

$$\langle Q_1(\mathbf{v}) \rangle_1 = \text{tr} \left\langle \mathbf{C}_1^{-1}(\mathbf{T}) \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_1 \right) \left(\mathbf{v} - \mathbf{T} \bar{\mathbf{g}}_1 \right)^\dagger \right\rangle_1 = L. \quad (\text{A1})$$

For the other expectation we need we have, after some computation,

$$\langle Q_2(\mathbf{v}) \rangle_1 = \text{tr} \left\{ \mathbf{C}_2^{-1}(\mathbf{T}) \left[\mathbf{C}_1(\mathbf{T}) + \mathbf{T} \mathbf{s} \mathbf{s}^\dagger \mathbf{T}^\dagger \right] \right\}. \quad (\text{A2})$$

Using properties of the trace we then have

$$2 \langle \lambda(\mathbf{v}) \rangle_1 = -L + \text{tr} \left[\mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{C}_1(\mathbf{T}) \right] + \mathbf{s}^\dagger \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{s} - \ln \det \left[\mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{C}_1(\mathbf{T}) \right].$$

This leads to the expression in the main text,

$$2F_1(\mathbf{T}) = -L + \text{tr} [\mathbf{R}_{12}(\mathbf{T})] + H_2(\mathbf{T}) - \ln \det [\mathbf{R}_{12}(\mathbf{T})]. \quad (\text{A3})$$

To compute $\nabla F_1(\mathbf{T})$ we start with

$$D_1(\mathbf{T}) = \frac{d}{dt} \text{tr} [\mathbf{R}_{12}(\mathbf{T} + t\mathbf{E})] \big|_{t=0}. \quad (\text{A4})$$

After some computation we find that $D_1(\mathbf{T})$ is given by

$$2 \text{tr} \left\{ \mathbf{E}^\dagger \left[-\mathbf{R}_{12}(\mathbf{T}) \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 + \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1 \right] \right\}. \quad (\text{A5})$$

Next, we have

$$D_2(\mathbf{T}) = \frac{d}{dt} \ln \det [\mathbf{R}_{12}(\mathbf{T} + t\mathbf{E})] \Big|_{t=0}. \quad (\text{A6})$$

A useful fact here is that

$$\frac{d}{dt} \ln \det [\mathbf{R}_{12}(\mathbf{T} + t\mathbf{E})] = \text{tr} \left[\mathbf{R}_{12}^{-1}(\mathbf{T} + t\mathbf{E}) \frac{d}{dt} \mathbf{R}_{12}(\mathbf{T} + t\mathbf{E}) \right]. \quad (\text{A7})$$

After some computation we find that $D_2(\mathbf{T})$ is given by

$$2 \text{tr} \left\{ \mathbf{E}^\dagger \left[-\mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 + \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1 \right] \right\}. \quad (\text{A8})$$

Finally, we have

$$D_3(\mathbf{T}) = \frac{d}{dt} H_2(\mathbf{T} + t\mathbf{E}) \Big|_{t=0}, \quad (\text{A9})$$

which eventually simplifies to

$$2 \text{tr} \left\{ \mathbf{E}^\dagger \left[-\mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{s} \mathbf{s}^\dagger \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 + \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{s} \mathbf{s}^\dagger \right] \right\}. \quad (\text{A10})$$

The gradient $\nabla F_1(\mathbf{T})$ is therefore

$$\mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \left(\mathbf{K}_1 + \mathbf{s} \mathbf{s}^\dagger \right) \left[\mathbf{I} - \mathbf{T}^\dagger \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 \right] - \mathbf{C}_2^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_2 + \mathbf{C}_1^{-1}(\mathbf{T}) \mathbf{T} \mathbf{K}_1. \quad (\text{A11})$$

APPENDIX B: AUC

We calculate the following derivatives first:

$$\frac{d}{dt} \mathbf{R}_{12}(\mathbf{T} + t\mathbf{E}) \Big|_{t=0} = \mathbf{C}_2^{-1}(\mathbf{T}) \left[-\left(\mathbf{E} \mathbf{K}_2 \mathbf{T}^\dagger + \mathbf{T} \mathbf{K}_2 \mathbf{E}^\dagger \right) \mathbf{R}_{12}(\mathbf{T}) + \left(\mathbf{E} \mathbf{K}_1 \mathbf{T}^\dagger + \mathbf{T} \mathbf{K}_1 \mathbf{E}^\dagger \right) \right], \quad (\text{B1})$$

and

$$\frac{d}{dt} \mathbf{R}_{21}(\mathbf{T} + t\mathbf{E}) \Big|_{t=0} = \mathbf{C}_1^{-1}(\mathbf{T}) \left[-\left(\mathbf{E} \mathbf{K}_1 \mathbf{T}^\dagger + \mathbf{T} \mathbf{K}_1 \mathbf{E}^\dagger \right) \mathbf{R}_{21}(\mathbf{T}) + \left(\mathbf{E} \mathbf{K}_2 \mathbf{T}^\dagger + \mathbf{T} \mathbf{K}_2 \mathbf{E}^\dagger \right) \right]. \quad (\text{B2})$$

The calculation of the required gradient is now straightforward, if a bit tedious. We define

$$\mathbf{Q}_1(\alpha, \mathbf{T}) = \mathbf{P}^{-1}(\alpha, \mathbf{T}) \mathbf{C}_1^{-1}(\mathbf{T}) = [\mathbf{C}_1(\mathbf{T}) \mathbf{P}(\alpha, \mathbf{T})]^{-1} \quad (\text{B3})$$

and

$$\mathbf{Q}_2(\alpha, \mathbf{T}) = \mathbf{P}^{-1}(\alpha, \mathbf{T}) \mathbf{C}_2^{-1}(\mathbf{T}) = [\mathbf{C}_2(\mathbf{T}) \mathbf{P}(\alpha, \mathbf{T})]^{-1}. \quad (\text{B4})$$

These are symmetric matrices. We also define

$$\mathbf{S}_1(\alpha, \mathbf{T}) = \mathbf{R}_{21}(\mathbf{T}) \mathbf{Q}_1(\alpha, \mathbf{T}) = [\mathbf{C}_1(\mathbf{T}) \mathbf{P}(\alpha, \mathbf{T}) \mathbf{R}_{12}(\mathbf{T})]^{-1} \quad (\text{B5})$$

and

$$\mathbf{S}_2(\alpha, \mathbf{T}) = \mathbf{R}_{12}(\mathbf{T}) \mathbf{Q}_2(\alpha, \mathbf{T}) = [\mathbf{C}_2(\mathbf{T}) \mathbf{P}(\alpha, \mathbf{T}) \mathbf{R}_{21}(\mathbf{T})]^{-1}. \quad (\text{B6})$$

These are also symmetric matrices. Then, the gradient of $P(\alpha, \mathbf{T})$ is $P(\alpha, \mathbf{T})$ times the matrix function

$$2 \left(\frac{1}{4} + \alpha^2 \right) \{ [\mathbf{Q}_2(\alpha, \mathbf{T}) - \mathbf{S}_1(\alpha, \mathbf{T})] \mathbf{T} \mathbf{K}_1 + [\mathbf{Q}_1(\alpha, \mathbf{T}) - \mathbf{S}_2(\alpha, \mathbf{T})] \mathbf{T} \mathbf{K}_2 \}. \quad (\text{B7})$$

Now define

$$\mathbf{U}_1(\mathbf{T}) = \int_{-\infty}^{\infty} [P(\alpha, \mathbf{T})]^{-\frac{1}{2}} [\mathbf{Q}_2(\alpha, \mathbf{T}) - \mathbf{S}_1(\alpha, \mathbf{T})] d\alpha, \quad (\text{B8})$$

and

$$\mathbf{U}_2(\mathbf{T}) = \int_{-\infty}^{\infty} [P(\alpha, \mathbf{T})]^{-\frac{1}{2}} [\mathbf{Q}_1(\alpha, \mathbf{T}) - \mathbf{S}_2(\alpha, \mathbf{T})] d\alpha. \quad (\text{B9})$$

Then, setting the gradient of $\text{AUC}(\mathbf{T})$ equal to zero is equivalent to

$$\mathbf{U}_1(\mathbf{T}) \mathbf{T} \mathbf{K}_1 = -\mathbf{U}_2(\mathbf{T}) \mathbf{T} \mathbf{K}_2. \quad (\text{B10})$$

We will assume that $\mathbf{U}_1(\mathbf{T})$ is invertible and check this assumption in the proof of Proposition 4. Then we have the familiar equation

$$\mathbf{K}_2^{-1} \mathbf{K}_1 \mathbf{T}^\dagger = -\mathbf{T}^\dagger \mathbf{U}_2(\mathbf{T}) \mathbf{U}_1^{-1}(\mathbf{T}). \quad (\text{B11})$$

This means that

$$-\mathbf{U}_2(\mathbf{T}) \mathbf{U}_1^{-1}(\mathbf{T}) = \mathbf{R}_{12}(\mathbf{T}). \quad (\text{B12})$$

REFERENCES

1. Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A.* 1987; 4:2379–2394. [PubMed: 3430225]
2. Barrett, H.; Myers, K. *Foundations of Image Science*. Wiley; 2013.

3. Abbey CK, Barrett HH, Eckstein MP. Practical issues and methodology in assessment of image quality using model observers. *Proc. SPIE*. 1997; 3032:182–194.
4. Kupinski M, Clarkson E, Hesterman J. Bias in Hotelling observer performance computed from finite data. *Proc. SPIE*. 2007; 6515:65150S.
5. Gallas BD. Variance of the channelized-hotelling observer from a finite number of trainers and testers. *Proc. SPIE*. 2003; 5034:100–111.
6. Nguyen N, Abbey C, Insana M. Objective assessment of sonographic quality I: Task information. *IEEE Trans. Med. Imaging*. 2013; 32:683–690. [PubMed: 23247846]
7. Abbey C, Zemp R, Liu J, Lindfors K, Insana M. Observer efficiency in discrimination tasks simulating malignant and benign breast lesions imaged with ultrasound. *IEEE Trans. Med. Imaging*. 2006; 25:198–209. [PubMed: 16468454]
8. Nguyen N, Abbey C, Insana M. Objective assessment of sonographic quality II: Acquisition information spectrum. *IEEE Trans. Med. Imaging*. 2013; 32:691–698. [PubMed: 23221818]
9. Kullback S, Leibler R. On information and sufficiency. *Ann. Math. Statistics*. 1951; 22:79–86.
10. Jeffreys H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. London A*. 1946; 186:453–461.
11. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin Cal. Math. Soc.* 1943; 35:99–109.
12. Barrett H, Abbey C, Clarkson E. Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions. *J. Opt. Soc. Am. A*. 1998; 15:1520–1535.
13. Peterson W, Birdsall T, Fox W. The theory of signal detectability. *Trans. IRE Prof. Group Inf. Theory*. 1954; 4:171–212.
14. Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J. Am. Coll. Radiol.* 2006; 3:413–422. [PubMed: 17412096]
15. Clarkson E. Asymptotic ideal observers and surrogate figures of merit for signal detection with list-mode data. *J. Opt. Soc. Am. A*. 2012; 29:2204–2216.
16. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J. Opt. Soc. Am. A*. 1987; 4:2447–2457. [PubMed: 3430229]
17. Park S, Kupinski MA, Clarkson E, Barrett H. Efficient channels for the ideal observer. *Proc. SPIE*. 2004; 5372:12–21.
18. Park S, Barrett H, Clarkson E, Kupinski M, Myers K. Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal. *J. Opt. Soc. Am. A*. 2007; 24:B136–B150.
19. Witten J, Park S, Myers K. Partial least squares: a method to estimate efficient channels for the ideal observers. *IEEE Trans. Med. Imaging*. 2010; 29:1050–1058. [PubMed: 20335088]
20. Park S, Witten J, Myers K. Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer. *IEEE Trans. Med. Imaging*. 2009; 28:657–668. [PubMed: 19272990]
21. Dollár P, Tu Z, Perona P, Belongie S. Integral channel features. *BMVC*. 2009; 2:5.
22. Barrett H, Yao J, Rolland J, Myers K. Model observers for assessment of image quality. *Proc. Natl. Acad. Sci. USA*. 1993; 90:9758–9765. [PubMed: 8234311]
23. Brankov J, Yang Y, Wei L, El-Naqa I, Wernick M. Learning a channelized observer for image quality assessment. *IEEE Trans. Med. Imaging*. 2009; 28:991–999. [PubMed: 19211351]
24. Field DJ. What is the goal of sensory coding? *Neural Computation*. 1994; 6:559–601.
25. Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*. 1998; 10:1299–1319.
26. Fukunaga K, Koontz WL. Application of the Karhunen-Loeve expansion to feature selection and ordering. *IEEE Trans. Comput.* 1970; C-19:311–318.
27. Huo X. A statistical analysis of Fukunaga-Koontz transform. *IEEE Signal Process. Lett.* 2004; 11:123–126.
28. Mahalanobis A, Muise RR, Stanfill SR, Van Nevel A. Design and application of quadratic correlation filters for target detection. *IEEE Trans. Aerosp. Electron. Syst.* 2004; 40:837–850.
29. Fukunaga, K. *Introduction to Statistical Pattern Recognition*. Academic; 1990.

30. Zhang S, Sim T. Discriminant subspace analysis: A Fukunaga-Koontz approach. *IEEE Trans Pattern Anal. Mach. Intell.* 2007; 29:1732–1745. [PubMed: 17699919]
31. Clarkson E, Shen F. Fisher information and surrogate figures of merit for the task-based assessment of image quality. *J. Opt. Soc. Am. A.* 2010; 27:2313–2326.
32. De la Torre, F.; Kanade, T. Proceedings of the 22nd International Conference on Machine Learning. ACM; 2005. Multimodal oriented discriminant analysis; p. 177-184.
33. Wimalajeewa T, Chen H, Varshney PK. Performance limits of compressive sensing-based signal classification. *IEEE Trans. Signal Process.* 2012; 60:2758–2770.
34. Carson WR, Chen M, Rodrigues MR, Calderbank R, Carin L. Communications-inspired projection design with application to compressive sensing. *SIAM J. Imag. Sci.* 2012; 5:1185–1212.
35. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory.* 2006; 52:489–509.
36. Donoho DL. Compressed sensing. *IEEE Trans. Inf. Theory.* 2006; 52:1289–1306.
37. Ji S, Xue Y, Carin L. Bayesian compressive sensing. *IEEE Trans. Signal Process.* 2008; 56:2346–2356.
38. Neifeld MA, Ashok A, Baheti PK. Task-specific information for imaging system analysis. *J. Opt. Soc. Am. A.* 2007; 24:B25–B41.
39. Ashok A, Baheti PK, Neifeld MA. Compressive imaging system design using task-specific information. *Appl. Opt.* 2008; 47:4457–4471. [PubMed: 18758516]
40. Davenport, MA.; Wakin, MB.; Baraniuk, RG. Technical Report TREE 0610. Department of ECE, Rice University; 2006. Detection and estimation with compressive measurements.
41. DasGupta, A. *Asymptotic Theory of Statistics and Probability.* Springer; 2008.
42. Clarkson E, Barrett H. Approximations to ideal-observer performance on signal-detection tasks. *Appl. Opt.* 2000; 39:1783–1793. [PubMed: 18345075]
43. Husemoller, D. *Fibre Bundles.* Springer; 1994.
44. Edelman A, Arias TA, Smith ST. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 1998; 20:303–353.
45. Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. Adaptation of retinal processing to image contrast and spatial scale. *Nature.* 1997; 386:69–73. [PubMed: 9052781]

Fig. 1.

Top image is a sample from the first class and the bottom image is a sample from the second class. The images look more similar in (a) since σ is smaller compared to (b). Movies for (a) and (b) are generated by repeatedly sampling from the ensemble (Media 1 and Media 2). (a) $\sigma = 0.2$ pixels. (b) $\sigma = 0.5$ pixels.

Fig. 2.

AUC comparison for (a) equal and (b) unequal mean images for the two classes. The AUC is calculated at various values of $\Delta\sigma$, the correlation length difference between the covariance matrices of the two classes. The number of channels is constant at $L = 25$ and the compression ratio $M/L = 100$. (a) Signal absent in the mean, (b) signal present in the mean.

Fig. 3.

Eigenspectrum of $\mathbf{K}_2^{-1}\mathbf{K}_1$ for six values of σ . The highest 25 values of $\kappa + 1/\kappa$ are marked in red (in the top left part of each curve) to denote the Eigen-CQO selections. The number of channels is constant at $L = 25$. $\sigma = 0.1$ pixels (a), 0.15 pixels (b), 0.2 pixels (c), 0.25 pixels (d), 0.30 pixels (e), 0.35 pixels (f).

Fig. 4.

Observer comparison for unequal mean images and varying the magnitude of the mean difference. The number of channels and second-order statistics are held constant; $L = 25$ and $\sigma = 0.15$ pixels.

(a)

(b)

Fig. 5. Observer comparison for (a) equal and (b) unequal mean images under both classes and varying the number of channels. The second-order statistics are constant; $\sigma = 0.15$ pixels. (a) Signal absent in the mean, (b) signal present in the mean.

Fig. 6.

J versus iteration number in Eq. (24) for (a) equal and (b) unequal means. Gradient term in Eq. (24) versus iteration number for (c) equal and (d) unequal means. The second-order statistics are constant; $\sigma = 0.25$ pixels. Discrepancies for different runs (using independent random samples for the initial solution) of the iterative gradient-based search algorithm indicate the presence of local maxima. (a), (c) Signal absent in the mean. (b), (d) Signal present in the mean.

Fig. 7.

Eigenspectrum of $\mathbf{R}_{12}$ [see Eq. (19)] for (a) equal and (b) unequal means and $\sigma = 0.25$ pixels and $L = 25$. Discrepancies for different runs (using independent random samples as the initial solution) of the iterative gradient-based search algorithm indicate the presence of local maxima. When a signal in the mean is introduced an eigenvalue close to one appears in the eigenspectrum. The corresponding channel is the matrix-vector product of $\mathbf{T}^\dagger$ and the corresponding eigenvector. This channel is shown in (c) to demonstrate that the signal in the mean affects the J-CQO channels. (a) Signal absent in the mean and (b) signal present in the mean. (c) $\mathbf{T}$.

Fig. 8.

Top row displays the channels (i.e., rows of the matrix $\mathbf{T}$) for the J-CQO iterative algorithm for three values of the correlation length difference between the two classes with equal means. $L = 25$ and each row of the channel matrix is displayed as a single image in a 5×5 tiling. The bottom row shows the eigenvectors of $\mathbf{R}_{12}$ [see Eq. (19)] back projected by $\mathbf{T}^\dagger$. (a) J-CQO $\mathbf{T}$ ($\sigma = 0.2$ pixels), (b) J-CQO $\mathbf{T}$ ($\sigma = 0.5$ pixels), (c) J-CQO $\mathbf{T}$ ($\sigma = 1$ pixel), (d) backprojection of $\mathbf{R}_{12}$ eigenvectors by $\mathbf{T}^\dagger$ ($\sigma = 0.2$ pixels), (e) backprojection of $\mathbf{R}_{12}$ eigenvectors by $\mathbf{T}^\dagger$ ($\sigma = 0.5$ pixels), (f) backprojection of $\mathbf{R}_{12}$ eigenvector by $\mathbf{T}^\dagger$ ($\sigma = 1$ pixel).