An Optimization Approach to Signal Extraction from Noisy Multivariate Data

T. Yokoo,*,1 B. W. Knight,† and L. Sirovich*

*Laboratory of Applied Mathematics, Mount Sinai School of Medicine, New York, New York 10029; and
†Laboratory of Biophysics, Rockefeller University, New York, New York 10021

We consider a problem of blind signal extraction from noisy multivariate data, in which each datum represents a system's response, observed under a particular experimental condition. Our prototype example is multipixel functional images of brain activity in response to a set of prescribed experimental stimuli. We present a novel multivariate analysis technique, which identifies the different activity patterns (signals) that are attributable to specific experimental conditions, without *a priori* knowledge about the signal or the noise characteristics. The extracted signals, which we term the *generalized indicator functions,* are optimal in the sense that they maximize a weighted difference between the signal variance and the noise variance. With an appropriate choice of the weighting parameter, the method returns a set of images whose signal-to-noise ratios satisfy some user-defined level of significance. We demonstrate the performance of our method in optical intrinsic signal imaging of cat cortical area 17. We find that the method performs effectively and robustly in all tested data, which include both real experimental data and numerically simulated data. The method of generalized indicator functions is related to canonical variate analysis, a multivariate analysis technique that directly solves for the maxima of the signal-to-noise ratio, but important theoretical and practical differences exist, which can make our method more appropriate in certain situations. © 2001 Academic Press

Key Words: image analysis; functional imaging; optical imaging; multivariate analysis; signal analysis.

INTRODUCTION

In many scientific experiments, the primary objective is to investigate the "cause and effect" relationship between the input and the output of a system. For example, in functional imaging of mammalian visual cortices, one observes the spatial patterns of brain response elicited by different experimental visual stimulations. Optical imaging (for a review, see Grinvald, 1992; Sirovich and Kaplan, 2001) is one such technique that measures the intrinsic optical signals from the cortex. These signals represent the activity-dependent changes in light absorption and scattering of the underlying tissue (Malonek and Grinvald, 1996; Malonek *et al.,* 1996; Frostig *et al.,* 1990; Arieli *et al.,* 1995). Optical imaging has contributed critical insights to how different submodalities of visual information are represented in V1 and V2 (primary and secondary visual areas), such as ocular dominance and orientation (Blasdel and Salma, 1986; Blasdel, 1992; Bonhoeffer and Grinvald, 1991), direction of motion (Weliky, 1996; Shmuel and Grinvald, 1996), and spatial frequency (Everson *et al.,* 1998; Issa *et al.,* 2000).

In optical imaging experiments, experimenters are typically interested in the different spatial patterns of brain activity under different functional states, which can be artificially created by controlling the experimental conditions. In the case of the visual cortex, the experimental conditions are often represented by different parameters of visual stimuli, such as orientation and spatial and temporal frequency. In our discussions, however, the conditions could be defined to be any kind of variable, continuous or categorical, including thermal or metabolic challenges to the brain tissue, various pharmacological treatments, or different disease states. In general, the images acquired under a single experimental condition can exhibit considerable frame-to-frame variations, despite the uniformly maintained experimental parameters. When the sources of these variations are not known or cannot be controlled, one might treat the observed image as a random variable, having some probability distribution. Often, it is assumed that the nonspecific response that is uncorrelated with the condition is the *noise.* In such cases, the condition-specific response is taken to be the difference

1 To whom correspondence should be addressed at A20-76, Laboratory of Applied Mathematics, Mount Sinai School of Medicine, One Gustav L. Levy Place, New York, NY 10029. Fax: (212) 426-5037. E-mail: yokoot01@doc.mssm.edu.

between the means of the two distributions, one from condition "A" and the other from condition "B".

By far the most popular estimator of the true mean is the "sample average," computed from repeated sampling of the probability distribution. The Law of Large Numbers then states that, in the limit of indefinitely many samples, the sample average approaches the true mean of the distribution. In actual experiments, however, the number of samples is limited, and the sample average is only an estimate, whose standard error is approximately proportional to $\sigma/\sqrt{N}$, where σ is the standard deviation of the response and N is the number of samples. This relation suggests that the error of the sample average can be made arbitrarily small by choosing an appropriate N. However, the required number of samples may be unrealistically large if the measurements come from a probability distribution with a large variance. Indeed in a typical optical imaging experiment, the condition-specific changes in activity can be significantly smaller than the background activity of the cortex, which fluctuates due to its intrinsic physiological processes, such as respiration, circulation, and autonomous neuronal activities (Arieli *et al.,* 1996). Since the signals must be reliably extracted from data containing exceedingly large background noise, analysis of optical imaging data presents a challenging problem.

A number of data analysis techniques have been devised to improve the signal recovery beyond that of the "difference image" (Orbach and Cohen, 1983), a pixel-by-pixel difference between two sample average images. These improved techniques include principal component analysis-based methods (Sirovich and Everson, 1992; Sirovich *et al.,* 1996), indicator function (Everson *et al.,* 1998), truncated difference (Gabbay *et al.,* 2000), independent component analysis (Bell and Sejnowski, 1995; Hyvarinen and Oja, 1996), and extended spatial decorrelation (Stetter *et al.,* 2000; Schiessl *et al.,* 2000). Here we present a new procedure, the method of *generalized indicator functions,* which extracts an orthogonal set of functions (images) that indicate the presence of condition-specific signals. In the next section, we will derive the method as a optimization problem of a statistic, a weighted *signal–noise difference* between the estimated signal variance and the noise variance. The method is related to the canonical discriminant function (Boch, 1975), which optimizes the signal-to-noise ratio, defined as a quotient of the signal variance and the noise variance. Although the techniques are related, there are important theoretical and practical differences, which will be illustrated and discussed in detail in later sections.

GENERALIZED INDICATOR FUNCTIONS

Motivation

The generalized indicator functions can be motivated quite generally for any multivariate data, but in the interest of concreteness, we will derive the method in the context of image analysis. Consider an optical imaging experiment, in which a charge-coupled device camera captures digital images of the cortex as it responds to different visual stimuli (for a review see Grinvald, 1992). Each image is a two-dimensional array of P pixels, whose pixel values are proportional to the number of photons arriving at those pixel locations on the camera's photodetector. Let $f_m(t, \mathbf{x})$ represent the pixel gray-level value at the cortical position $\mathbf{x} = (x, y)$, in the tth frame under experimental condition m, where $t = 1, \ldots, T$, and $m = 1, \ldots, M$, and T is the same for each m. For fixed t and m, and the full range of $\mathbf{x}$, $f_m(t, \mathbf{x})$ is a single image, and it can be regarded as a vector whose components are indexed by $\mathbf{x}$. In our discussion we often regard the pixel index $\mathbf{x}$ as implicit and refer to the image as $f_m(t)$.

We define $\bar{f}_m$, the average image for the mth condition, and $\bar{f}$, the average image over all conditions, as

$$\bar{f}_m = \frac{1}{T} \sum_{t=1}^{T} f_m(t) \tag{1}$$

and

$$\bar{f} = \frac{1}{M} \sum_{m=1}^{M} \bar{f}_m. \tag{2}$$

Then, $\bar{f}_m - \bar{f}$ is an estimate of the mean image under condition m, measured from a reference image, which we choose to be $\bar{f}$. Similarly, $f_m(t) - \bar{f}_m$ is the variation of the individual images not explained by $\bar{f}_m$.

We now seek an as yet unknown *image,* $\phi(\mathbf{x})$, with which we hope to estimate the map of the condition-specific response, the signal map. Since we have no *a priori* knowledge of the spatial or temporal characteristics of such a signal, the only objective criterion for signal identification is the presence of a strong correlation between the signal and the experimental conditions. We will exploit this relationship and require the "strength" of the signal image to vary as a function of the experimental condition. Such an image, $\phi(\mathbf{x})$, which we term the *generalized indicator function,* indicates the presence of a signal within the dataset. We formalize this argument by introducing the *response amplitude function* $\rho_m(t)$ of the response map $\phi(\mathbf{x})$.

With the inner product between two vectors defined as the standard Euclidean dot product, consider the projections onto $\phi(\mathbf{x})$

$$\rho_m(t) = (f_m(t), \phi)_{\mathbf{x}}, \tag{3}$$

where the subscript $\mathbf{x}$ indicates that the inner products are formed by the sum over the pixels. The inner product $\rho_m(t)$ is a functional of ϕ and represents the degree

to which $\phi(\mathbf{x})$ is present in $f_m(t, \mathbf{x})$. Hence, for any fixed image $\phi(\mathbf{x})$, $\rho_m(t)$ represents the amplitude of $\phi(\mathbf{x})$ within the tth data image under the mth experimental condition, where T frames are taken from each of M conditions. If we define

$$c = (\bar{f}, \phi)_{\mathbf{x}},$$

$$d_m = (\bar{f}_m - \bar{f}, \phi)_{\mathbf{x}}, \qquad (4)$$

$$e_m(t) = (f_m(t) - \bar{f}_m, \phi)_{\mathbf{x}},$$

the response amplitude function can be decomposed into three parts:

$$\rho_m(t) = c + d_m + e_m(t). \qquad (5)$$

We may reasonably refer to c as the background amplitude, d_m as the signal amplitude, and $e_m(t)$ as the noise amplitude. The desired $\phi(\mathbf{x})$ is then a function which has maximal variations in d_m between experimental conditions $m = 1, \ldots, M$, while having minimal residual variations in $e_m(t)$ for any fixed condition m.

The Generalized Indicator Function Problem

We proceed to construct a quantitative criterion for identification for the desired signal maps. The sample variance of the signal amplitude can be measured by the *mean sum of squares between conditions,*

$$S(\phi) = \frac{1}{M-1} \sum_m \{d_m\}^2$$

$$= \frac{1}{M-1} \sum_m \{(\bar{f}_m - \bar{f}, \phi)_{\mathbf{x}}\}^2, \qquad (6)$$

and it can be identified as the average signal power per image. If we define the *signal operator* as

$$K_S(\mathbf{x}, \mathbf{y}) = \frac{1}{M-1} \sum_m (\bar{f}_m(\mathbf{x}) - \bar{f}(\mathbf{x}))(\bar{f}_m(\mathbf{y}) - \bar{f}(\mathbf{y})), \quad (7)$$

then

$$S(\phi) = (\phi, K_S\phi)_{\mathbf{x}}. \qquad (8)$$

Similarly, the sample variance of the noise amplitude can be measured by the *mean sum of squares within conditions,*

$$N(\phi) = \frac{1}{M(T-1)} \sum_{m,t} \{e_m(t)\}^2$$

$$= \frac{1}{M(T-1)} \sum_{m,t} \{(f_m - \bar{f}_m, \phi)_{\mathbf{x}}\}^2, \qquad (9)$$

and it is identified as the average noise power per image. If we define the *noise operator* as

$$K_N(\mathbf{x}, \mathbf{y}) = \frac{1}{M(T-1)} \sum_{m,t} (f_m(t, \mathbf{x}) - \bar{f}_m(\mathbf{x}))$$

$$\times (f_m(t, \mathbf{y}) - \bar{f}_m(\mathbf{y})), \qquad (10)$$

then

$$N(\phi) = (\phi, K_N\phi)_{\mathbf{x}}. \qquad (11)$$

Note that both K_S and K_N are real, symmetric, and positive semidefinite operators. Thus their eigenvalues are real and nonnegative, and each has a set of eigenvectors which are real, orthogonal, and complete.

In order to quantify the success of $\phi(\mathbf{x})$, we define a performance measure, which we call the *signal–noise difference,*

$$S(\phi) - \alpha N(\phi), \qquad (12)$$

where the scalar weighting parameter α is yet to be specified. It is necessary to apply a constraint on the length of $\phi(\mathbf{x})$, which we take to be

$$\|\phi\|_{\mathbf{x}}^2 = (\phi, \phi)_{\mathbf{x}} = 1. \qquad (13)$$

It follows that the criterion function $C(\phi)$ for this constrained maximization problem is

$$C(\phi) = S(\phi) - \alpha N(\phi) - \gamma\|\phi\|_{\mathbf{x}}^2, \qquad (14)$$

or

$$C(\phi) = (\phi, \{K_S - \alpha K_N - \gamma\}\phi)_{\mathbf{x}}, \qquad (15)$$

where γ is a Lagrange multiplier.

In order to find the extremum of $C(\phi)$, we denote the desired extremum by ϕ_o and consider admissible variations,

$$\phi(\mathbf{x}; \epsilon) = \phi_o(\mathbf{x}) + \epsilon\Delta(\mathbf{x}), \qquad (16)$$

for an arbitrary $\Delta(\mathbf{x})$. The stationarity condition then states:

$$\frac{dC(\epsilon)}{d\epsilon}\bigg|_{\epsilon=0} = 0. \qquad (17)$$

We apply this to Eq. (15) and obtain the eigenvalue problem,

$$(K_S - \alpha K_N)\phi = \gamma\phi. \qquad (18)$$

Since both K_S and K_N are real symmetric $P \times P$ operators, where P is the number of pixels per image, the operator in Eq. (18) is also a symmetric $P \times P$ operator, regardless of the choice of α. Therefore, this eigenvalue problem yields a total of P eigenpairs, $\{\gamma_k, \phi_k\}$, $k = 1, \ldots, P$, and the eigenvectors can be taken to be a complete orthonormal set with real eigenvalues.

Interpreting the Solutions

Solutions to the generalized indicator function problem Eq. (18) are the extrema of the signal–noise difference, $S(\phi) - \alpha N(\phi)$, under the unit-length constraint, and further analysis is necessary to identify the subset of the solutions that correspond to the maxima. In this section, we present a physical interpretation of the eigenvalue problem and the role of the weighting constant α as a quality-control parameter for obtaining the solutions of desirable quality.

Recall that our strategy for signal identification is through the behavior of the signal in different experimental conditions. If a signal image is found, its behavior is characterized by the condition-correlated changes in its amplitude. We have seen that, for a fixed image $\phi(\mathbf{x})$, its response amplitude function can be regarded as a combination of the background, signal, and noise components,

$$\rho_m(t) = c + d_m + e_m(t). \qquad (19)$$

Since the signal variance, $S(\phi)$, and the noise variance, $N(\phi)$, can be identified with the power of the signal amplitude and the power of the noise amplitude, respectively, the quotient

$$Q(\phi) = \frac{S(\phi)}{N(\phi)} = \frac{(\phi, K_S\phi)_{\mathbf{x}}}{(\phi, K_N\phi)_{\mathbf{x}}} \qquad (20)$$

can be regarded as a measure of signal-to-noise ratio of the image $\phi(\mathbf{x})$.

In order to provide a convincing argument that the kth eigenvector, $\phi_k(\mathbf{x})$, represents a reliable map of the condition-specific response, its amplitude function must have a sufficiently large signal-to-noise ratio. We devise a systematic approach to test a candidate image by setting a critical Q level, q_c, above which we accept it as the signal of interest. Thus, q_c represents the minimum acceptable signal-to-noise ratio. Throughout this paper, we use the minimum acceptable signal-to-noise of $q_c = 4$, that is, we consider an image to be a signal if its mean amplitude changes by at least two standard deviations of the noise.

Now observe, from Eqs. (7) and (10), that solutions for the generalized indicator problem satisfy the property

$$(\phi_k, \{K_S - \alpha K_N\}\phi_k)_{\mathbf{x}} = \gamma_k \|\phi_k\|_{\mathbf{x}}^2 \qquad (21)$$

and because of the constraint, $\|\phi_k\|_{\mathbf{x}}^2 = 1$, we have,

$$S(\phi_k) - \alpha N(\phi_k) = \gamma_k. \qquad (22)$$

By rearranging the above equation, and using the non-negativity of $N(\phi)$, we have an inequality for all positive γ_k,

$$Q(\phi_k) = \frac{S(\phi_k)}{N(\phi_k)} = \alpha + \frac{\gamma_k}{N(\phi_k)} > \alpha; \text{ for all } \gamma_k > 0. \qquad (23)$$

This relation suggests that, if α is chosen to be $\alpha = q_c$, each eigenvector ϕ_k associated with a positive eigenvalue is guaranteed to have desirable Q level. Moreover, it can be shown that all linear combinations of such eigenvectors satisfy the minimum signal-to-noise requirement, and thus the orthonormal basis set $\{\phi_k; \gamma_k > 0\}$ forms a vector subspace of $Q > q_c$. The weighting constant α $(=q_c)$ thus functions as a quality control parameter for the generalized indicator problem. Because of this feature, the maximization of $S(\phi) - \alpha N(\phi)$ can be viewed as a procedure that seeks out images with the maximal signal power, $S(\phi)$, within a subspace whose members have sufficiently large signal-to-noise ratios.

It is also worth noting that the signal-to-noise ratio $Q(\phi)$, Eq. (20), is defined as a ratio of variances, and it is related to Fisher's F in statistics,

$$F(\phi) = TQ(\phi) = \frac{TS(\phi)}{N(\phi)}, \qquad (24)$$

where T is the number of frames per condition. Following the conventional statistical approach for significance tests, one may suppose a null hypothesis that the amplitude of $\phi(\mathbf{x})$ is normally distributed and that its mean amplitude does not change due to different conditions, i.e., there is no signal. Under the null hypothesis, the F statistic is a random variable having an F distribution with $M - 1$ and $M(T - 1)$ degrees of freedom. This reference distribution provides a means to quantify the level of confidence with which we can conclude that a signal truly exists within the data. For our chosen minimum signal-to-noise requirement ($q_c =$

4), the F values of the images in $Span\{\phi_k; \gamma_k > 0\}$ correspond to confidence levels of virtually 100%.

Computation of the Eigenvectors

The generalized indicator functions can be obtained by solving the eigenvalue equation (18). In actual situations, however, the number of pixels, P, can be as large as $O(10^5)$, which makes the direct pixel-to-pixel solution of the $P \times P$ problem impractical. Fortunately, a simple observation, which has led to the *snapshot method* (Sirovich, 1987; Sirovich and Everson, 1992), shows that a smaller yet still exact calculation is possible.

In particular, we note that the operator in Eq. (18) is the weighted sum of two operators, defined by Eqs. (7) and (10). Hence, the action of these operators on $\phi(\mathbf{x})$ always results in some linear superposition of the elements of $\{\bar{f}_m - \bar{f}\}$ and $\{f_m(t) - \bar{f}_m\}$. Any member of $\{\bar{f}_m - \bar{f}\}$ or $\{f_m(t) - \bar{f}_m\}$ must be expandable as a linear combination of the mean-corrected images, $\{f_m(t, \mathbf{x}) - \bar{f}\}$, since

$$\bar{f}_m - \bar{f} = \frac{1}{T} \sum_t (f_m(t) - \bar{f}) \qquad (25)$$

and

$$f_m(t) - \bar{f}_m = (f_m(t) - \bar{f}) - (\bar{f}_m - \bar{f}). \qquad (26)$$

This implies that $\phi(\mathbf{x})$ is necessarily an admixture of the mean-corrected images of the data, $\{f_m(t, \mathbf{x}) - \bar{f}\}$. Since there are MT frames (snapshots) in the dataset, $\phi(\mathbf{x})$ can be represented as a linear combination of no more than MT such images. The direct application of this observation, although correct and practical, does not offer the most felicitous framework. Instead, we consider a more refined formulation based on principal component analysis (PCA), which also illuminates the underlying mathematical structure.

To this end, we follow Sirovich and Everson (1992) and seek to write every mean-corrected image, $f_m(t, \mathbf{x}) - \bar{f}$, in the form

$$f_m(t, \mathbf{x}) - \bar{f} = \sum_{r=1}^{MT} a_m^r(t) \sigma_r \psi_r(\mathbf{x}) \qquad (27)$$

such that

$$(a_m^i(t), a_m^j(t))_{m,t} = \delta_{i,j},$$
$$(\psi_i, \psi_j)_{\mathbf{x}} = \delta_{i,j}. \qquad (28)$$

The above series expansion is analogous to a procedure known as singular value decomposition (SVD) in matrix algebra, and in our discussions PCA and SVD may be considered equivalent.

Now, if the above biorthogonality requirements are applied to Eq. (27),

$$(a_m^r(t), f_m(t, \mathbf{x}) - \bar{f}(\mathbf{x}))_{m,t} = \sigma_r \psi_r(\mathbf{x}) \qquad (29)$$

and

$$(\psi_r, f_m(t, \mathbf{x}) - \bar{f}(\mathbf{x}))_{\mathbf{x}} = \sigma_r a_m^r(t) \qquad (30)$$

must hold. If we now use our data to define the two operators,

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{MT} \sum_{m,t} (f_m(t, \mathbf{x}) - \bar{f}(\mathbf{x}))$$
$$\times (f_m(t, \mathbf{y}) - \bar{f}(\mathbf{y})), \qquad (31)$$

$$C(m, t, n, s) = \frac{1}{MT} \sum_{\mathbf{x}} (f_m(t, \mathbf{x}) - \bar{f}(\mathbf{x}))$$
$$\times (f_n(s, \mathbf{y}) - \bar{f}(\mathbf{y})), \qquad (32)$$

then back substitution of Eq. (29) into (30), and similarly of Eq. (30) into (29), leads to the two eigenvalue problems,

$$\sum_{\mathbf{y}} K(\mathbf{x}, \mathbf{y}) \psi_r(\mathbf{y}) = \lambda_r \psi_r(\mathbf{x}), \qquad (33)$$

$$\sum_{n,s} C(m, t, n, s) a_n^r(s) = \lambda_r a_m^r(t), \qquad (34)$$

where the eigenvalue $\lambda_r = \sigma_r^2$.

Because the operators K and C are both symmetric, the corresponding eigenvectors are orthogonal and complete in their respective spaces. The expansion of the form (27) is now implemented. The method of snapshots takes advantage of the fact that Eq. (34) is a much smaller problem than (33) when $P \gg MT$. After Eq. (34) is solved, the spatial eigenvectors $\{\psi_r(\mathbf{x})\}$ can be constructed from Eq. (29).

Once given a decomposition of the form (27), we can rewrite $f_m(t) - \bar{f}_m$ and $\bar{f}_m - \bar{f}$ as

$$f_m(t) - \bar{f}_m = \sum_{r=1}^{MT} \{a_m^r(t) - \langle a_m^r(t) \rangle_t\} \sigma_r \psi_r$$
$$= \sum_{r=1}^{MT} \{a_m^r(t) - \bar{a}_m^r\} \sigma_r \psi_r \qquad (35)$$

and

$$\bar{f}_m - \bar{f} = \sum_{r=1}^{MT} \langle a_m^r(t) \rangle_t \sigma_r \psi_r = \sum_{r=1}^{MT} \bar{a}_m^r \sigma_r \psi_r. \qquad (36)$$

The earlier remark that $\phi(\mathbf{x})$ is a linear superposition of $\{f_m(t, \mathbf{x}) - \bar{f}\}$ can be taken one step further, and we may write $\phi(\mathbf{x})$ as a linear combination in terms of Eq. (27),

$$\phi(\mathbf{x}) = \sum_{r=1}^{MT} b_r \psi_r(\mathbf{x}), \qquad (37)$$

where b_r is the yet unknown coefficient on the rth spatial eigenvector.

If we now substitute Eqs. (35), (36), and (37) into the generalized indicator function problem (18), and use the orthonormal property of the ψ_r's, we obtain an equivalent $MT \times MT$ eigenvalue problem for the generalized indicator functions,

$$D(C_S - \alpha C_N)D\mathbf{b} = \gamma\mathbf{b}, \qquad (38)$$

where

$$\mathbf{b} = [b_1, b_2, \ldots, b_{MT}]^{\dagger}, \qquad (39)$$

$$C_S(i, j) = \frac{1}{MT} \sum_{m,t} (\bar{a}_m^i)(\bar{a}_m^j), \qquad (40)$$

$$C_N(i, j) = \frac{1}{MT} \sum_{m,t} (a_m^i(t) - \bar{a}_m^i)(a_m^j(t) - \bar{a}_m^j), \qquad (41)$$

$$D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{MT} \end{bmatrix}. \qquad (42)$$

The $\dagger$ symbol denotes the transpose of a vector or a matrix. The operator of Eq. (38) is again real and symmetric, so the eigenvectors are guaranteed to be orthogonal and complete, with real eigenvalues. When the snapshot-space dimension, MT, is smaller than the pixel-space dimension, P, this alternative formulation is the computationally smaller problem to solve. Once the solutions for Eq. (38) are obtained, we can easily determine $\phi(\mathbf{x})$ from Eq. (37).

OTHER ANALYSIS METHODS

The Introduction mentions several analysis methods for signal identification in highly variable data. In order to establish a frame of reference for the generalized indicator method, we discuss two alternatives here. First is the differential imaging procedure, which is the most common analysis method for optical imaging data. Another analysis technique, called canonical variate analysis, also deserves attention, because of its kinship to the method of generalized indicator functions. It has its origin in multigroup classification and statistical decision theory, but can also be applied to

signal extraction problems in a natural way as demonstrated by Friston *et al.* (1995).

Differential Imaging

In differential imaging, the sample average of the nth condition is subtracted from the sample average of the mth condition to produce a difference image $\Delta_{m,n}(\mathbf{x})$:

$$\Delta_{m,n}(\mathbf{x}) = \bar{f}_m(\mathbf{x}) - \bar{f}_n(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} f_m(t, \mathbf{x}) - \frac{1}{T}\sum_{t=1}^{T} f_n(t, \mathbf{x}). \qquad (43)$$

In the limit indefinitely large and unbiased samples ($T \to \infty$), each sample mean approaches its true mean, and then $\Delta_{m,n}(\mathbf{x})$ represents the true signal difference between conditions m and n. When the condition n is taken to be some "control" condition, $\bar{f}_n(\mathbf{x})$ estimates the condition-independent activity, and its subtraction from $\bar{f}_m(\mathbf{x})$ is essentially an attempt to remove the background. The resulting difference image estimates the characteristic activity map under the condition m. However, with a limited number of samples T, the standard error of this estimate at pixel location $\mathbf{x}$ is approximately $\sigma_{\mathbf{x}}/\sqrt{T}$, where $\sigma_{\mathbf{x}}$ is the standard deviation of the pixel value at position $\mathbf{x}$. Hence, the errors of the sample averages may significantly corrupt the differential image, in the event of a large variance in the response probability distribution.

Canonical Variate Analysis

Under Generalized Indicator Functions we motivated our analysis method as a maximization problem of the signal–noise difference and then argued that the desired solutions are the ones whose signal-to-noise ratios exceed some threshold constant q_c. Instead, one may opt to look for a function $\chi(\mathbf{x})$ that directly maximizes the ratio

$$Q(\chi) = \frac{S(\chi)}{N(\chi)} = \frac{(\chi, K_S\chi)_{\mathbf{x}}}{(\chi, K_N\chi)_{\mathbf{x}}}. \qquad (44)$$

The maximization of a quotient of two quadratic forms, called a Rayleigh quotient, is a classical problem (Bellman, 1960), and this particular form of the Rayleigh quotient has been treated extensively in multivariate statistics under several aliases, such as canonical variate analysis, canonical discriminant analysis, and multiple discriminant analysis. The maximization of the ratio (44) can be shown to be equivalent to a generalized eigenvalue problem (Boch, 1975),

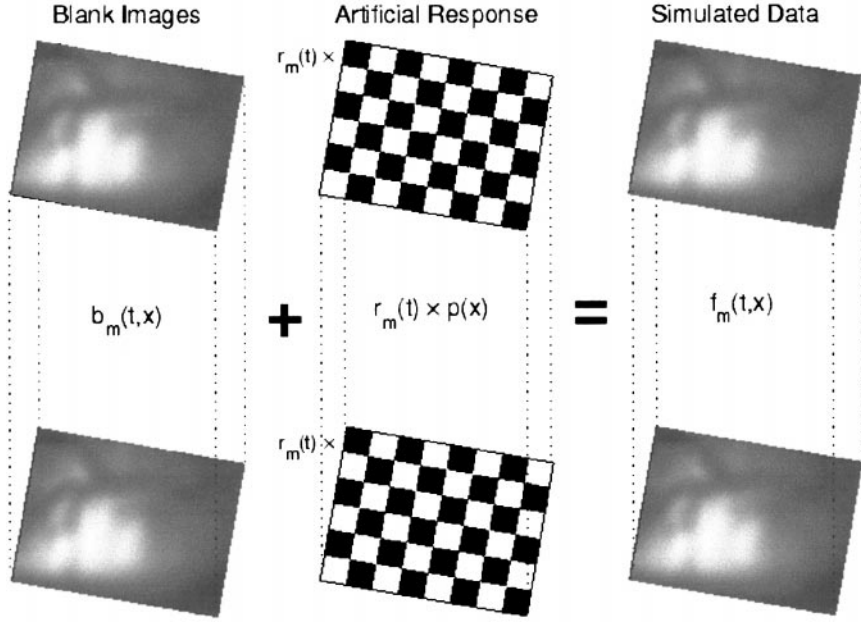$$K_S\chi_k = \lambda_k K_N\chi_k. \qquad (45)$$

FIG. 1. Construction of the simulated dataset. (Left) A set of normalized "blank" images, $b_m(t, \mathbf{x})$; $\|b(t, \mathbf{x})\|_\mathbf{x}^2 = 1$, of cat primary visual cortex. (Middle) The artificial signal is modeled as a checkerboard pattern $p(\mathbf{x})$, $\|p(\mathbf{x})\|_\mathbf{x}^2 = 1$, whose amplitude in each data image is determined by $r_m(t)$. (Right) The simulated dataset is generated by adding the artificial signal to the blank set of images. Since $r_m(t)$ is typically chosen to be $O(10^{-3}) - O(10^{-4})$ for all m and t, the checkerboard pattern $p(\mathbf{x})$ cannot be visually detected in the data images.

The eigenvector χ_k is known as the kth canonical variate and satisfies the following property,

$$\lambda_k = \frac{(\chi_k, \, K_S\chi_k)_\mathbf{x}}{(\chi_k, \, K_N\chi_k)_\mathbf{x}}. \tag{46}$$

Thus, when ordered in descending magnitude, the eigenvalue λ_k is the kth largest local maximum of the signal-to-noise ratio, associated with the canonical variate χ_k. Thus, canonical variate analysis seeks the images $\{\chi_k(\mathbf{x})\}$ such that the power of the signal, $(\chi_k, \, K_S\chi_k)_\mathbf{x}$, is maximized *relative* to the power of the noise, $(\chi_k, \, K_N\chi_k)_\mathbf{x}$. We also note that the maxima of Q also correspond to the maxima of the F ratio, as these quantities are proportional to one another. For this reason, the canonical variate analysis is often presented in the literature as a maximization problem of the F ratio.

In practice, we choose to solve a different form of the same problem, instead of solving Eq. (45) directly. This alternative formulation, which we derive in part A of the Appendix, transforms Eq. (45) into a standard eigenvalue problem, for which computationally efficient algorithms exist.

NUMERICAL SIMULATIONS

Numerically simulated data will be used to illustrate the method of generalized indicator functions and to compare its performance with differential imaging and canonical variate analysis. Artificial data will be constructed from a known signal and well-characterized background noise. Each technique will attempt to recover the signal from the noisy data, and their solutions will be quantitatively evaluated, as the true signal is known.

An overview of the data construction is shown in Fig. 1. We construct the data, a collection of 96×128 pixel images, by adding an artificial spatial pattern $p(\mathbf{x})$ to a set of 660 unstimulated, or "blank," images of cat primary visual cortex, $b_m(t, \mathbf{x})$. These images were acquired while the anesthetized cat was viewing a blank, uniformly illuminated screen, thus assumed to contain no stimulus-evoked signals. We design the artificial response map as a "checkerboard" pattern $p(\mathbf{x})$, $\|p(\mathbf{x})\|_\mathbf{x}^2 = 1$. In order to simulate the correlated changes in the response amplitude due to stimuli, an artificial amplitude function $r_m(t)$ is introduced, so that the response signal is modeled as

$$s_m(t, \mathbf{x}) = r_m(t)p(\mathbf{x}). \tag{47}$$

We do not specify the amplitude function $r_m(t)$ now, but once we choose a specific $r_m(t)$, the synthetic data $f_m(t, \mathbf{x})$ can be constructed by adding the model signal to the blank images,

$$f_m(t, \mathbf{x}) = b_m(t, \mathbf{x}) + s_m(t, \mathbf{x}), \tag{48}$$

where the images in $b_m(t, \mathbf{x})$ are normalized; $\|b_m(t)\|_\mathbf{x}^2 = 1$. The amplitude function $r_m(t)$ thus determines the

amount of $p(\mathbf{x})$ added to each of the blank images. In all simulations that follow, the signal will be added at $O(10^{-3}) - O(10^{-4})$ of the background strength, so the pattern $p(\mathbf{x})$ cannot be perceived by the naked eye.

For the purpose of simulation, we "prepare" the background, so that it is orthogonal to the signal,

$$(b_m(t), \, p)_{\mathbf{x}} = 0, \tag{49}$$

hence, $p(\mathbf{x})$ is projected out of the blank set. It therefore follows that

$$(f_m(t), \, p)_{\mathbf{x}} = r_m(t). \tag{50}$$

The purpose of this step will become clear in the next two sections, when we consider situations in which the background is *not* orthogonal to the signal.

We consider a two-group experiment in which $M = 2$ and $T = 330$. The first 330 frames of $f_m(t, \mathbf{x})$ are assigned to condition group $m = 1$ and the remaining 330 to group $m = 2$. The condition-specific responses are then given by $r_1(t)\,p(\mathbf{x})$ and $r_2(t)\,p(\mathbf{x})$, respectively.

Now, the goal of our analysis is to recover an estimated signal $\phi(\mathbf{x})$ from the dataset, $f_m(t, \mathbf{x})$, in the absence of any information about $p(\mathbf{x})$ or $r_m(t)$. If the recovery were perfect, we would have $\phi = p$, and thus the amplitude $\rho_m(t)$ of the recovered signal should equal the modeled amplitude $r_m(t)$,

$$\rho_m(t) = (f_m(t), \, \phi)_{\mathbf{x}} = (f_m(t), \, p)_{\mathbf{x}} = r_m(t). \tag{51}$$

We can quantify the success of signal recovery by $\phi(\mathbf{x})$ since the true signal $p(\mathbf{x})$ is known. Since ϕ and p are unit vectors, we have $\cos \theta = (\phi, \, p)_{\mathbf{x}}$, where θ is the angle between ϕ and p vectors. Then the efficiency defined as

$$E = (\phi, \, p)_{\mathbf{x}}^2 \tag{52}$$

gives unity when $\phi(\mathbf{x}) = p(\mathbf{x})$ and zero when $\phi(\mathbf{x}) \perp p(\mathbf{x})$. We also examine the signal-to-noise ratio of the recovered signal, $Q(\phi)$, in comparison with the model signal's signal-to-noise ratio $Q(p)$.

A Nonstochastic Signal

We first model the response signal whose amplitude changes only as a function of the condition index m with no frame-to-frame variation in t. Let the artificial amplitude function be

$$r_m(t) = \begin{cases} -5 \times 10^{-4} & \text{if } m = 1 \\ +5 \times 10^{-4} & \text{if } m = 2 \end{cases} \tag{53}$$

as shown in Fig. 2A. The weights $\pm 5 \times 10^{-4}$ are chosen so that signal is added to the blank images at a signal-to-background ratio of 5:10,000, a typical
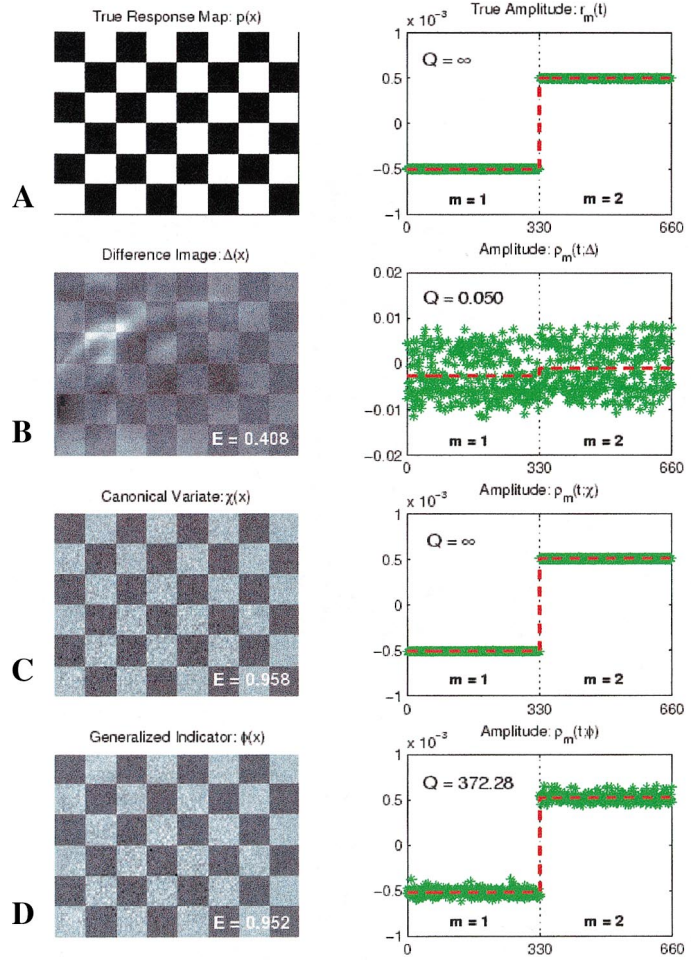

FIG. 2. A two-condition simulated experiment with a nonstochastic signal. (A) The model signal pattern and its amplitude used in the simulation. The estimated signal and its amplitude extracted by (B) differential imaging, (C) canonical variate analysis, and (D) the method of generalized indicator functions are shown.

strength of the intrinsic signals in optical imaging. Since $r_m(t)$ has zero variability in t, so does $\rho_m(t)$ by Eq. (50), and the signal-to-noise ratio of $p(\mathbf{x})$ becomes unbounded.

We first examine the result from the differential imaging procedure (Fig. 2B). A difference image $\Delta(\mathbf{x})$ is computed as $\Delta_{1,2}(\mathbf{x})$ using Eq. (43), and its response amplitude is given by $\rho_m(t; \Delta) = (f_m(t), \Delta)_{\mathbf{x}}$. Although $\Delta(\mathbf{x})$ shows a definite resemblance to the true signal ($E = 0.408$), it is corrupted by a significant amount of noise from vascular pulsations, vasomotor effects, and other less obvious sources. Since $\Delta(\mathbf{x})$ carries the expected signal within it, one might be misled to believe that $\Delta(\mathbf{x})$ is a good representation of a real difference between conditions $m = 1$ and $m = 2$. However, from a signal-vs-noise point of view, such a statement neglects the fact that the response amplitude $\rho_m(t; \Delta)$ shows extensive within-condition variability. The small signal-to-noise ratio, $Q(\Delta) = 0.0508$, suggests that the

underlying signal is obliterated by the presence of excessive noise.

The canonical variate analysis guarantees the optimal signal-to-noise ratio. The solution $\chi(\mathbf{x})$ and its amplitude $\rho_m(t; \chi) = (f_m(t), \chi)_\mathbf{x}$ are displayed in Fig. 2C, which shows highly successful signal recovery with $E = 0.958$ and a faithful reproduction of $r_m(t)$. The calculated signal-to-noise ratio of χ is greater than the machine-infinity (eps^{-1}) and we denote it $Q(\chi) = \infty$. This extreme signal-to-noise ratio is in agreement with the completely noise-free nature of the modeled response amplitude function, $r_m(t)$. Why and how the canonical analysis can achieve an unbounded signal-to-noise ratio is explained in part B of the Appendix.

The generalized indicator $\phi(\mathbf{x})$ (Fig. 2D) is computed with $q_c = 4$, which sets the minimum acceptable signal-to-noise ratio. The quality of the solution, $\phi(\mathbf{x})$, is comparable to that of the canonical variate, as it recovers $E = 0.952$ of the original signal. The response amplitude $\rho_m(t; \phi) = (f_m(t), \phi)_\mathbf{x}$ offers a highly significant signal-to-noise ratio, $Q(\phi) = 372.28$, although it does not achieve the greatest possible signal-to-noise ratio, as demonstrated by the canonical variate. The amplitude of $\rho_m(t; \phi)$ is distributed tightly around $\pm 5 \times 10^{-4}$ and gives a good approximation of the modeled response amplitude, $r_m(t)$.

A Stochastic Signal

Next, we consider data in which the signal of interest not only changes in amplitude as a function of the condition index m, but also fluctuates randomly in the frame number t. Let $r_m(t)$ be

$$r_m(t) = \begin{cases} -5 \times 10^{-4} \cdot \text{IIDN}(1, \sigma^2) & \text{if } m = 1 \\ +5 \times 10^{-4} \cdot \text{IIDN}(1, \sigma^2) & \text{if } m = 2, \end{cases} \quad (54)$$

where $\text{IIDN}(1, \sigma^2)$ denotes an independently, identically, and normally distributed random variable with mean of 1, and here we let the noise variance be $\sigma^2 = 1/4$ (Fig. 3A). Therefore, $r_1(t)$ and $r_2(t)$ are samples from normal distributions centered around $\pm 5 \times 10^{-4}$. Using the definition (20) and a little arithmetic, the expectation of $Q(p)$ can be shown to be 8. Since we use $q_c = 4$ in these simulations, this artificial signal, despite the presence of noise, is considered significant. This addition of the stochastic component is our attempt to model two equivalent situations: (i) the signal itself is intrinsically noisy or (ii) the signal itself is nonstochastic, but the background already contains randomly fluctuating image components that, regarded as vectors, are parallel to the true signal. As far as the analysis is concerned, these two situations are mathematically indistinguishable, because the noise sources are assumed to be unknown.

The result from the differential imaging procedure (Fig. 3B) is qualitatively identical to that from the
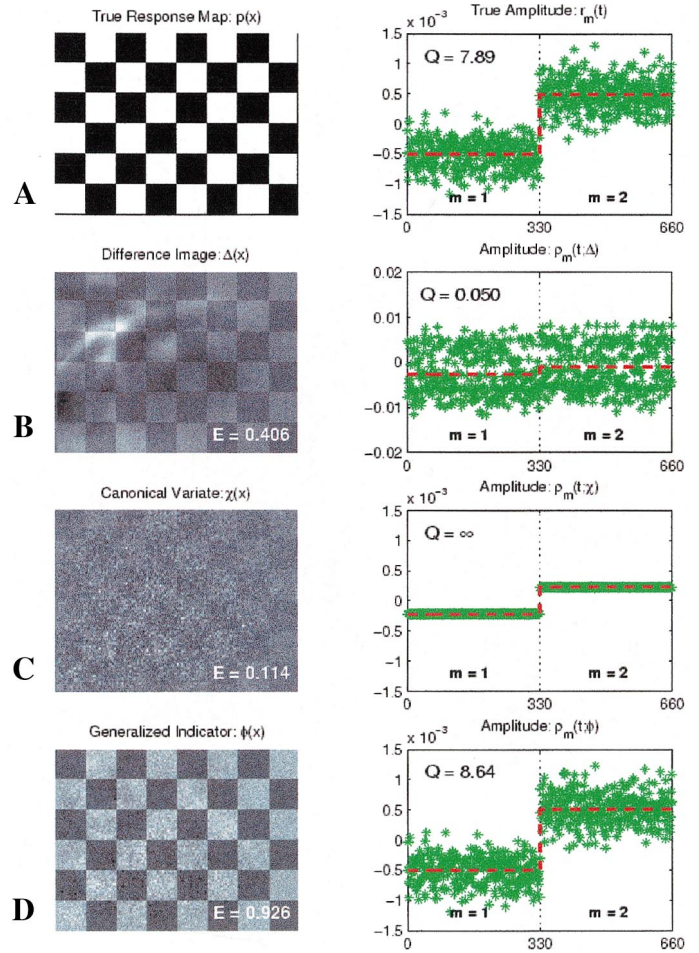

FIG. 3. A two-condition simulated experiment with a stochastic signal. (A) The model signal pattern and its amplitude used in the simulation. The estimated signal and its amplitude, extracted by (B) differential imaging, (C) canonical variate analysis, and (D) generalized indicator function, are shown.

previously considered noise-free case. Significant distortion by the background noise is present in $\Delta(\mathbf{x})$, which shows only a modest signal recovery, $E = 0.406$, and the amplitude $\rho_m(t, \Delta)$ has extensive within-group variation, which results in a small signal-to-noise ratio, $Q(\Delta) = 0.050$.

The most striking consequence of this randomly fluctuating signal is the drastic quality deterioration in the canonical variate $\chi(\mathbf{x})$ (Fig. 3C). The solution $\chi(\mathbf{x})$ shows poor signal recovery with $E = 0.114$; the image is heavily infiltrated by pixel noise, and the spatial pattern $p(\mathbf{x})$ is hardly discernible. On the other hand, its amplitude function $\rho_m(t; \chi) = (f_m(t), \chi)_\mathbf{x}$ shows the optimal resolution with $Q(\chi) = \infty$, even though the true signal has only $Q(p) \approx 8$. It is also evident from the plot of $\rho_m(t; \chi)$ that the amplitude of the recovered signal is considerably attenuated, compared to the modeled response amplitude. We see that canonical analysis has overestimated the signal-to-noise ratio, by choosing a
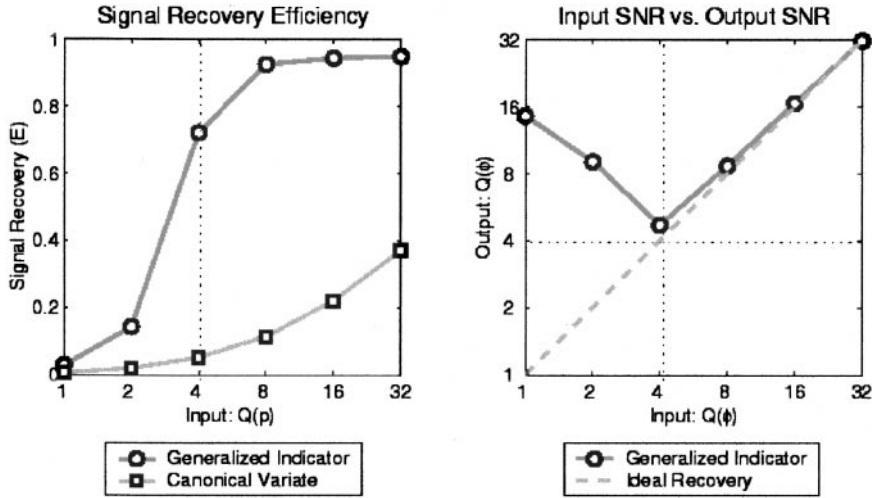
FIG. 4. Noise analyses. The variance of the Gaussian fluctuations in the artificial signal amplitude, $r_m(t)$, is modulated such that signal-to-noise ratio of the true signal $p(\mathbf{x})$ corresponds to $Q(p) = 1, 2, 4, 8, 16, 32$. (Left) A plot of the signal recovery efficiency, $E = (p, \phi)_{\mathbf{x}}^2$, as a function of $F(p)$. (Right) A plot of the recovered signal's Q value as a function of the true signal's Q value.

spurious pattern which does not contain any frame-to-frame noise.

Figure 3D shows the result from the analysis with the generalized indicator method. Despite the introduction of noise, the solution $\phi(\mathbf{x})$ maintains excellent signal recovery at $E = 0.926$. The recovered response amplitude $\rho_m(t; \phi) = (f_m(t), \phi)_{\mathbf{x}}$ preserves a significant signal-to-noise ratio at $Q(\phi) = 8.647$, which is close to the true signal's signal-to-noise ratio, $Q(p) \approx 8$ (see Fig. 3A). It also recaptures the stochastic nature of the modeled response amplitude $r_m(t)$, with no overall amplitude attenuation. The robustness of the generalized indicator method against noise follows from its more reasonable demand on the signal, namely, $Q(p) > q_c$, which permits its realistic applications to stochastic signals with suboptimal signal-to-noise ratio.

Noise Analyses

Similar noise analyses can be performed with the addition of Gaussian noise at various strengths. The noise variance σ^2 of the stochastic signal (Eq. (54)) has been modulated to be $\sigma^2 = 2, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$, and the simulation and the analysis of the previous section is repeated with these progressively less noisy signals. Figure 4 (left) shows the plot of signal recovery E as a function of the corresponding signal-to-noise ratio of the model signal $Q(p) = 1, 2, 4, 8, 16, 32$. When $Q(p)$ is below the critical value $q_c = 4$, both methods show poor signal recovery. As $Q(p)$ is increased past the critical value, the generalized indicator rapidly outperforms the canonical variate and reaches its maximum recovery at about $E = 0.95$. On the other hand, the signal recovery of the canonical variate improves only sluggishly as the true signal is made less noisy. Figure 4 (right) shows the relationship between the signal-to-

noise ratio of the model signal $Q(p)$ and that of the generalized indicator, $Q(\phi)$. If the recovery were perfect, we would have $Q(\phi) = Q(p)$, and the recovery curve would fall on the diagonal line. However, when the model signal is exceedingly noisy such that $Q(p) < q_c$, the generalized indicator tends to overestimate the signal-to-noise ratio, as the algorithm always guarantees $Q(\phi) > q_c$. Such overestimation error in $Q(\phi)$ is consistent with the poor signal recovery performance by $\phi(\mathbf{x})$ in this low signal-to-noise regime. On the other hand, when $Q(p) > q_c$, the true signal becomes statistically significant, and the generalized indicator's $Q(\phi)$ closely approximates the signal-to-noise ratio of the true response. Incidentally, the recovered signal-to-noise ratio of the canonical variate is consistently unbounded, regardless of the amount of noise introduced into the model signal (for explanation, see part B of Appendix). This result suggests the relative robustness of the generalized indicator method over the canonical variate analysis in the signal extraction problem of noisy, yet reasonably robust, signals.

APPLICATION TO EXPERIMENTAL DATA

In this section we reexamine the experimental data (Everson *et al.,* 1998) which have been considered previously by two techniques, truncated difference (Gabbay *et al.,* 2000) and indicator function (Everson *et al.,* 1997). We will briefly review these techniques and point out how the generalized indicator function method improves upon them, in terms of analytic simplicity, generality, and objectivity. We see that, when applied to actual experimental data, our method produces qualitatively similar results, with better statistical characteristics and less subjective intervention.

Previous Analysis Strategies

The method of Gabbay *et al.* (2000), which may be regarded as a refinement of the standard differences (see Differential Imaging), is based on the observation that the difference image (Eq. (43)) is expected to converge to the true response map in the limit of indefinitely many samples. When the sample size is limited, principal component analysis and a statistical analysis of the principal modes make possible the identification of the modal index range for the most relevant and significant principal components; from this an improved difference image, called the *truncated difference,* can be constructed. On the other hand, the indicator function method of Everson *et al.* (1997) is, in spirit, akin to Fisher's discriminant function (Fisher, 1936). As with the truncated difference method, it involves a band-pass filtering of the principal component spectrum. Although the construction of such filters is warranted by sound principles in both methods, an element of subjectivity enters into the analyses through the investigator's choice of the pass-band. In contrast, no explicit filter is introduced into the generalized indicator procedure; instead, the selection of signal-associated principal components is performed implicitly by the optimization algorithm. Another distinction is that these earlier methods are inherently one-dimensional procedures, as they generate a single image (vector) by contrasting two experimental conditions at one time. For experiments that use multiple conditions ($M > 2$), the response space can be M dimensional. To deal with such cases, Everson *et al.* (1998) introduce a "pairwise" analysis strategy and compute multiple contrast images for all possible pairs of conditions. A subsequent principal component analysis on all pairwise contrast images then determines the basis elements of the response space. Although such a pairwise procedure makes good mathematical sense, it is unnecessary in the generalized indicator method, as the procedure directly addresses the multidimensional nature of the signal space.

Data Analyses

In the experiment of Everson *et al.* (1998), a cat is presented with a visual stimulus, a drifting sinusoidal grating oriented at 0, 45, 90, and 135°, at six different spatial frequencies (0.07, 0.14, 0.28, 0.57, 1.14, and 2.28 cycles/deg), and the cortical intrinsic signals are imaged. For each orientation–spatial frequency combination, 40 frames are imaged, for a total of 960 frames, with 240 frames per orientation. Here we analyze the orientation-specific response by pooling all the images corresponding to each orientation. We set m so that $\theta = [0, 45, 90, 135]$ corresponds to $m = [1, 2, 3, 4]$. The following analyses are performed strictly on raw data, and we do not introduce any *ad hoc* processing of the
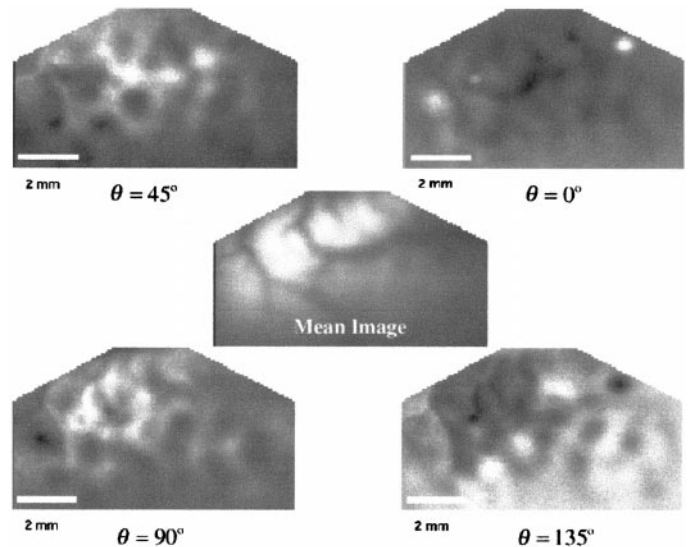

FIG. 5. An analysis of actual optical imaging data. A cat is presented with a series of sinusoidal grating, oriented at 0, 45, 90, and 135°, and 240 images of the primary visual cortex are imaged for each orientation, i.e., $M = 4$ and $T = 240$. The cortex is illuminated with 600-nm light. The mean image over all orientations, $\bar{f}$, which we take as the reference image, is shown in the center. The difference image, $\bar{f}_m - \bar{f}$, is also shown for each orientation.

data, such as whitening, spatial/temporal filters, or any kinds of normalization or approximation.

The four sample mean images for each orientation are shown in Fig. 5. Although no obvious differences can be visually detected, the stimulus-specific signal is presumably present in each image. Figure 6 shows the two dominant basis functions (images) and their respective response amplitudes for the orientation response space, extracted by: (Fig. 6A) pairwise truncated differences, (6B) pairwise indicator functions, (6C) canonical variates, and (6D) generalized indicator functions (computed with $q_c = 4$). These pairs of basis images represent the four estimates of the "signal space," each calculated by different algorithms. For each analysis method, the estimated neural response to any particular orientation can be expressed as a linear combination of the two basis images.

The basis images obtained by the pairwise truncated differences show superior smoothness and clarity of the emergent patterns, but their response amplitudes show large within-group variations, and their signal-to-noise ratios are $Q(\tau_1) = 1.472$ and $Q(\tau_2) = 1.158$, respectively. The pairwise indicator basis images are strikingly similar to the corresponding generalized indicator functions and their amplitudes have better signal-to-noise ratios, $Q(\iota_1) = 3.775$ and $Q(\iota_2) = 2.982$. The canonical variates are heavily contaminated by pixel noise, although some underlying structures are visible. A careful comparison reveals that the canonical variates actually represent extremely noisy versions of the same maps as the generalized indicator functions. The
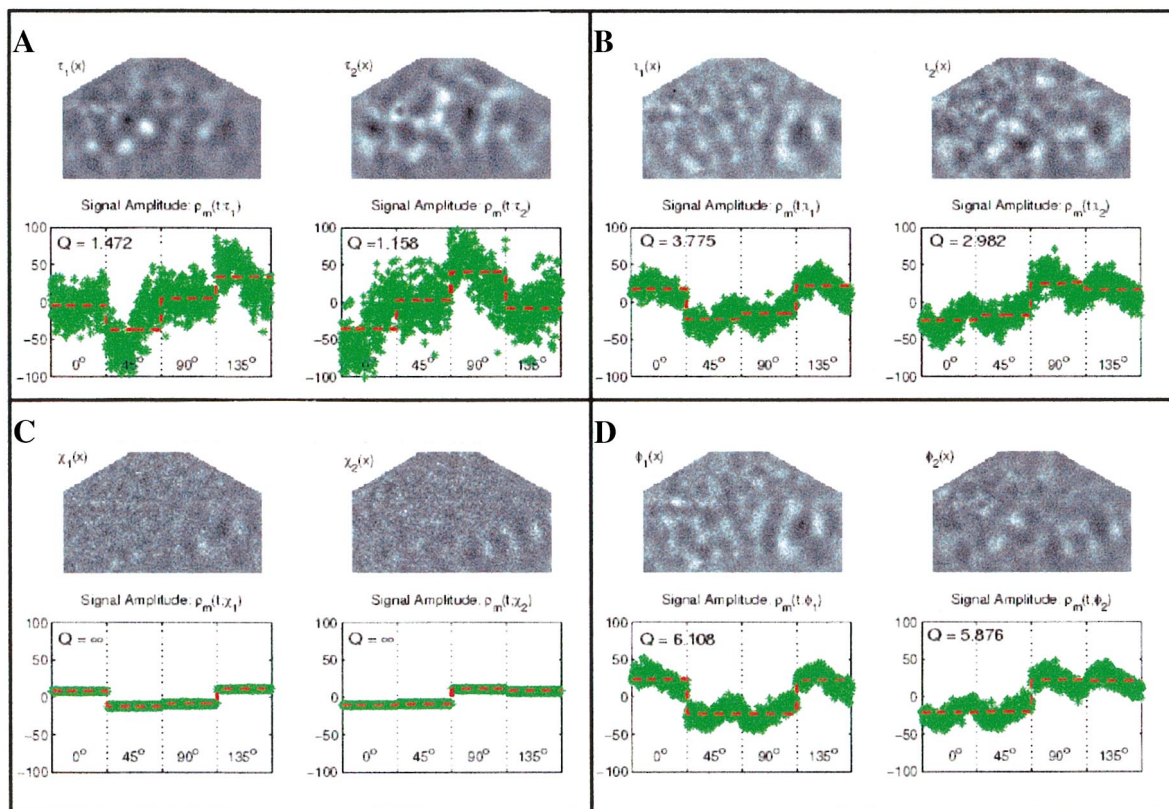
FIG. 6. Two dominant basis functions of the orientation response space and their corresponding amplitude functions, extracted by: (A) pairwise truncated difference method, (B) pairwise indicator function method, (C) canonical variate analysis, and (D) generalized indicator function method.

unbounded signal-to-noise ratios of their response amplitudes are consistent with the objective of the canonical analysis and demonstrate that such extreme signal-to-noise ratio can be achieved within the limits of the data. The generalized indicator function algorithm guarantees that the solutions, if found, always yield acceptable signal-to-noise ratio. Indeed, the first two generalized indicators exceed the minimum signal-to-noise requirement, $Q(\phi_1) = 6.108$ and $Q(\phi_2) = 5.878$, respectively. Note that any method could potentially return basis images whose signal-to-noise ratios clear the minimum requirement level. However, the generalized indicator functions are unique among all images of significant signal-to-noise ratios, because they represent the images of the maximal signal power, and thus their amplitude changes most robustly as a function of orientation. One remarkable feature seen in this analysis is the regular repeating pattern in amplitude functions. Recall that there is nothing in the algorithm that forces these patterns to emerge. The systematic variation within each oriented stimulus presentation is presumably due to the different response strength to different spatial frequencies, which we do not address here.

From the two generalized indicator functions and their respective amplitude functions, an orientation preference map can be constructed (for method, see Everson *et al.,* 1998). Each colored pixel value of such maps represents the stimulus orientation at which the pixel's stimulus-evoked response is maximal. Figure 7 shows the orientation preference map obtained from the generalized indicators, after a spatial smoothing by a low-pass filter. The map crisply shows the pinwheel structure of the orientation columns, indicated by asterisks, in agreement with previous reports (Blasdel and Salma, 1986; Bonhoeffer and Grinvald, 1991; Everson *et al.,* 1998). These results show that the generalized indicator method can be useful in the analysis of actual experimental data, as it generates solutions that are not only desirable in their statistical properties, but also consistent with the results from other analysis techniques and/or from other laboratories.

DISCUSSION

Generalized Indicator and Canonical Variates

The relationship between generalized indicator functions and canonical variates deserves further discussion. These analysis techniques are similar in the sense that they both use, directly or indirectly, the signal-to-noise ratio as a guide to identify the signals
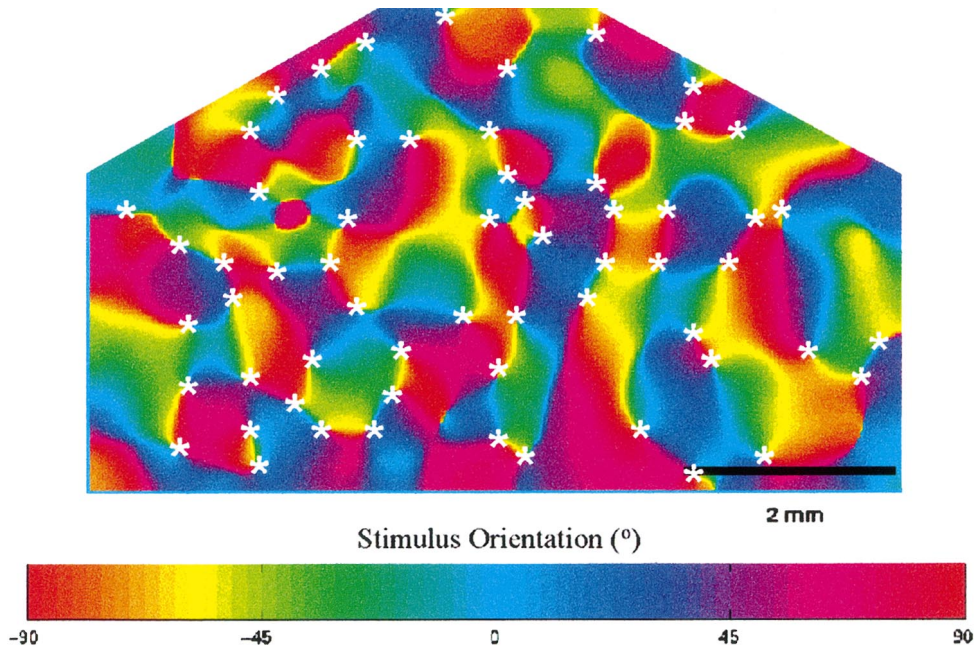
Stimulus Orientation (°)

-90   -45   0   45   90

FIG. 7. An orientation preference map is constructed from the two generalized indicator functions, $\phi_1(\mathbf{x})$ and $\phi_2(\mathbf{x})$, and their respective amplitudes. The color in each pixel represents the stimulus orientation at which the greatest stimulus-specific response is evoked. The map reveals a clear pinwheel pattern on the cortex (white asterisks), and the preferred angle varies smoothly around each pinwheel center.

that are hidden in the background noise. Their fundamental difference lies in their requirements for the extraction of such signals. Loosely speaking, the canonical variate analysis aims to extract the "cleanest" signals, while the generalized indicator method seeks the strongest signals that are "clean enough."

From a signal analysis point of view, the ratio of the two variances, $Q = S/N$, represents the magnitude of the signal power, $S(\phi)$, *relative* to the power of the noise, $N(\phi)$. By design, the canonical variate analysis maximizes the signal-to-noise ratio and does so irrespective of the absolute magnitude of the signal. Hence, it is possible for the solutions to achieve the maxima while compromising the absolute signal power, as seen in the example under Numerical Simulations. On the other hand, the method of generalized indicators attempts to maximize the signal–noise difference, $S(\phi) - \alpha N(\phi)$, a performance measure that respects the absolute power of the signal. Because the weighting parameter α is chosen such that the solution's signal-to-noise ratio is always significant, the method can be understood as a technique which seeks the most "robust" signal, while preserving our ability to discriminate a real causal relationship from random fluctuations. Hence, the method of generalized indicator functions may be used as an alternative to canonical variate analysis when one wishes to focus on the *absolute* difference rather than the *relative* difference between the power of the signal and that of the noise.

A pictorial illustration of the problem's subspace structures is presented in part B of the Appendix, as it facilitates the understanding of the similarities and the differences between the two analysis methods.

Robustness

We have subjected a wider variety of optical imaging data to the generalized indicator method (than reported here), and we have found that it performs robustly in all instances. When applied to the data such as those discussed under Application to Experimental Data, the new method consistently returns results that are qualitatively similar to the previous analysis results, but with superior sensitivity and signal–noise properties. Application to more recent data, such as those from a thermal and GABA-mediated V2 inactivation experiment (Sailstad, 2001) and a V1 color discrimination experiment (Gegiu *et al.,* 2001), gives reliable results that are consistent among different experimental animals. Our experience suggests that the method is robust and well-behaved in optical imaging data, and we anticipate that it will be useful in other imaging media, such as fMRI and PET.

Comments on the Assumptions

Although the method of the generalized indicator functions has been presented here as an image analysis technique, the algorithm itself is very general, as it

can be applied to practically any type of multivariate data. However, a meaningful interpretation of the solutions may require the following assumptions to hold: (i) the data contain no spurious group-dependent bias and (ii) the signal-to-noise ratio of the true response is at least q_c. If these assumptions fail to hold, the analysis becomes less effective in capturing the true response and leads to compromised signal extraction efficiency.

The presence of a spurious group-dependent bias (Assumption i) could be a serious challenge to the "scientific" validity of the generalized indicator. This can arise because of the change in the experimental environment or the change in the basic physiological states of the animal, such as depth of anesthesia, CO_2 level, blood pressure. Since the analysis extracts the largest variations among different experimental conditions (within the confines of $Q > q_c$), regardless of the source of the variations, the presence of such bias can contaminate the extracted signals. Eliminating such bias is an important issue in the experimental design, and well-randomized data acquisition sequences could keep such undesirable effects at a minimum.

Assumption ii states that the success of the analysis by the generalized indicator functions depends on the success of the experiment itself. We saw under Noise Analyses that the generalized indicator method showed poor signal recovery when the true response's signal-to-noise ratio was below the chosen value of minimum Q level. The method is predicated on the basis that $Q(\phi) > q_c$, and it is theoretically impossible for the method to reliably recover the hidden signals whose signal-to-noise ratios are below q_c. Thus the data must contain a robust response with sufficient signal-to-noise ratio, and it is a prerequisite for a valid signal identification.

CONCLUSION

We have considered the problem of extracting signals from multivariate data that are highly contaminated by noise from unknown sources. We have presented a novel data analysis technique, the method of *generalized indicator functions,* which identifies the image components of the noisy data, which represent the characteristic response differences between experimental conditions. The method of generalized indicator functions is related to a multivariate analysis technique called the canonical variate analysis, which aims to directly maximize the signal-to-noise ratio. The fundamental difference between the two methods is that the canonical variate analysis' sole objective is the maximization of the signal-to-noise ratio, irrespective of actual magnitude of the signal power, while the method of generalized indicator functions attempts to achieve the maximum signal power, while maintaining sufficiently large signal-to-noise ratios. Unlike many

other existing techniques, the algorithm does not require any explicit filtering procedures, nor does it depend on statistical or probabilistic assumptions about the data. In the application to actual and numerically simulated experimental data, we have demonstrated excellent performance and robustness of generalized indicator functions, in comparison with the standard difference method and the canonical variate analysis. Since this method may be applied to any vectorizable data type, it may improve the signal extraction in different kinds of experiments in which the conventional analysis methods suffer from insufficient signal-to-noise ratio.

APPENDIX

(A) Computation of Canonical Variates

A computationally efficient form of canonical variate analysis is derived here as a transformation of a generalized eigenvalue problem into a regular eigenvalue problem of lower dimension. Although not necessary, for the sake of notational simplicity we annotate the signal and the noise operators as

$$K_S(\mathbf{x}, \mathbf{y}) = \frac{1}{M} \sum_m (\bar{f}_m(\mathbf{x}) - \bar{f}(\mathbf{x}))(\bar{f}_m(\mathbf{y}) - \bar{f}(\mathbf{y})), \quad (55)$$

$$K_N(\mathbf{x}, \mathbf{y}) = \frac{1}{MT} \sum_{m,t} (f_m(t, \mathbf{x}) - \bar{f}_m(\mathbf{x}))(f_m(t, \mathbf{y}) - \bar{f}_m(\mathbf{y})). \quad (56)$$

The only difference from the previous definitions (7) and (10) is in the leading normalization constants, which do not alter the validity of the following arguments. Recall that the canonical variate $\chi(\mathbf{x})$ is a solution to a maximization problem for the Q ratio,

$$Q(\chi) = \frac{(\chi, K_S\chi)_{\mathbf{x}}}{(\chi, K_N\chi)_{\mathbf{x}}}. \quad (57)$$

In this Appendix, we prefer to use the following equivalent notation,

$$Q(\chi) = \frac{\chi K_S\chi}{\chi K_N\chi}, \quad (58)$$

in which the inner products are written as matrix multiplications, e.g., $(\chi, K_S\chi)_{\mathbf{x}} = \chi K_S\chi$.

Now, consider the identity

$$K_S(\mathbf{x}, \mathbf{y}) + K_N(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}), \quad (59)$$

where $K(\mathbf{x}, \mathbf{y})$ is the mean-corrected data covariance operator, defined in Eq. (31). Then the Q ratio can be written as

$$Q(\chi) = \frac{\chi K_S \chi}{\chi K_N \chi} = \frac{\chi K \chi}{\chi K_N \chi} - 1. \qquad (60)$$

Hence, the extremization of Eq. (44) can be performed equivalently on the quotient,

$$Q'(\chi) = \frac{\chi K \chi}{\chi K_N \chi}. \qquad (61)$$

Note that this problem is analogous to Eq. (45), hence the solutions can be obtained by solving the generalized eigenvalue problem,

$$K\chi = \mu K_N \chi, \qquad (62)$$

where the eigenvalue μ represents the quotient Q' of the solution χ,

$$\mu = \frac{\chi K \chi}{\chi K_N \chi}. \qquad (63)$$

Note the eigenvalue μ has a lower bound, $\mu \geq 1$, because Eq. (60) is nonnegative.

Now, we make three observations which directly lead to the transformation of Eq. (62) into a regular eigenvalue problem.

Observation 1

The direct consequence of principal component analysis (Eq. (27)) is a spectral decomposition of K,

$$K = \Psi \Lambda \Psi^{\dagger}, \qquad (64)$$

where Λ is an $R \times R$ diagonal matrix of the real positive eigenvalues $\{\lambda_1, \ldots, \lambda_R\}$; $\lambda_1 \geq \cdots \geq \lambda_R > 0$; $\Psi = [\psi_1(\mathbf{x}), \ldots, \psi_R(\mathbf{x})]$ is a $P \times R$ matrix of corresponding orthonormal eigenvectors (the principal components of the mean-corrected images, $\{f_m(t, \mathbf{x}) - \bar{f}(\mathbf{x})\}$); and R is the number of nonzero eigenvalues. Here, without loss of generality, we can exclude from our calculations the zero eigenvalues and their corresponding eigenvectors, because they have no contribution in the matrix multiplication (Eq. (64)). If we exclude zero eigenvectors, Λ becomes invertible, because all diagonal elements are >0.

Observation 2

By the definition (56) of K_N, the right-hand side of Eq. (62) is necessarily a linear combination of $\{f_m(t) - \bar{f}_m\}$ and, therefore, of $\{f_m(t) - \bar{f}\}$ by Eq. (26). Similarly,

Eq. (62) yields a linear combination of $\{f_m(t) - \bar{f}\}$ on the left-hand side. This implies that the solution χ can be uniquely determined within and up to the subspace spanned by $\{f_m(t) - \bar{f}\}$. Therefore, in the spirit of principal component analysis (Eq. (27)), we write χ as

$$\chi(\mathbf{x}) = \sum_{r=1}^{R} w_r \psi_r(\mathbf{x}), \qquad (65)$$

or in a vector notation,

$$\chi = \Psi \mathbf{w}; \text{ where } \mathbf{w} = [w(1), \ldots, w(R)]^{\dagger}. \qquad (66)$$

Observation 3

Another consequence of principal component analysis (Eq. (27)) is that the K_N operator can be written as

$$K_N = \Psi D C_N D \Psi^{\dagger}, \qquad (67)$$

where C_N and D are defined in Eqs. (41) and (42), respectively.

Now the substitution of Eqs. (64), (66), and (67) into (62) yields the equation

$$\mu^{-1}(\Psi \Lambda \Psi^{\dagger}) \Psi \mathbf{w} = (\Psi D C_N D \Psi^{\dagger}) \Psi \mathbf{w}. \qquad (68)$$

The division by μ is allowed, because $\mu \geq 1$. (See Eq. (63).)

Because $\Psi^{\dagger} \Psi = \mathbf{I}$, where $\mathbf{I}$ is an $R \times R$ identity matrix, and $\Lambda = D^2$, Eq. (62) can be reduced to an $R \times R$ regular eigenvalue problem:

$$\mu^{-1} \mathbf{w} = (D^{-1} C_N D) \mathbf{w}. \qquad (69)$$

Because C_N and D are symmetric, the operator of Eq. (69) is real and symmetric, so the eigenvalues are real, and the eigenvectors $\mathbf{w}_k$, $k = 1, \ldots, R$ are orthonormal and complete in the R-dimensional space. The canonical variate χ can then be computed by the linear combination (Eq. (65)). Since the eigenvalue μ_k^{-1} is the reciprocal of the quotient $Q'(\chi_k)$, the eigenvector $\mathbf{w}_k$ associated with the kth smallest μ_k^{-1} determines the desired kth local maximum, χ_k. Hence, we have successfully transformed a $P \times P$ generalized eigenvalue problem (Eq. (62)) for the canonical variate analysis into a smaller $R \times R$ regular eigenvalue problem.

The canonical variates can be shown to be an orthogonal set. Because of Eq. (65) and orthonormality of $\{\psi_r(\mathbf{x})\}$ and $\{\mathbf{w}_k\}$ in their respective spaces,

$$(\chi_i, \chi_j)_{\mathbf{x}} = \sum_{\mathbf{x}} \{\sum_{r=1}^{R} w_i(r) \psi_r(\mathbf{x})\} \{\sum_{s=1}^{R} w_j(s) \psi_s(\mathbf{x})\} \qquad (70)$$
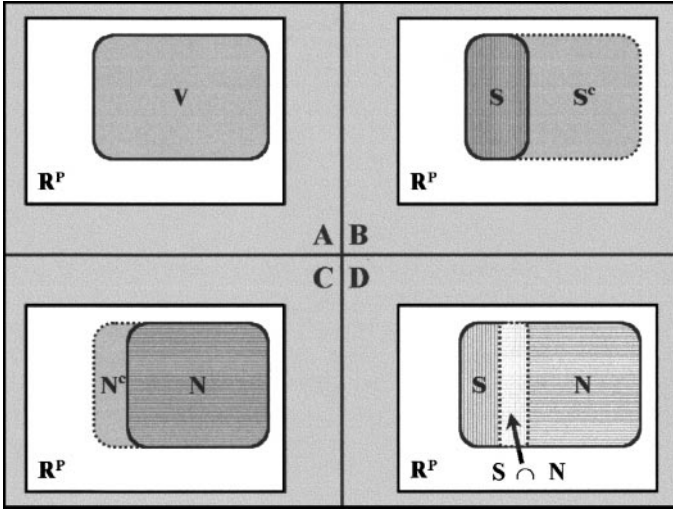$$= (\mathbf{w}_i, \mathbf{w}_j)_r = \delta_{i,j}.$$

FIG. 8. The subspace structure of the canonical variate and the generalized indicator problems.

(B) The Subspace Structures

Here we illustrate the subspace structures of the canonical variate and the generalized indicator problems (Fig. 8). We also make an informal argument why the canonical variates analysis leads to $Q(\chi) \to \infty$ in general.

To begin, we first note that the two analysis methods share common grounds; namely, both involve vector operations in the subspaces

$$\mathbf{V} \equiv \{\mathbf{v} \in \mathbf{R}^P; \; \mathbf{v}K\mathbf{v} > 0\} = \text{Range}(K),$$

$$\mathbf{S} \equiv \{\mathbf{v} \in \mathbf{R}^P; \; \mathbf{v}K_S\mathbf{v} > 0\} = \text{Range}(K_S), \quad (71)$$

$$\mathbf{N} \equiv \{\mathbf{v} \in \mathbf{R}^P; \; \mathbf{v}K_N\mathbf{v} > 0\} = \text{Range}(K_N),$$

where K, K_S, and K_N are $P \times P$ matrix operators defined in Eqs. (31), (55), and (56), respectively (Figs. 8A, 8B, and 8C). By construction, the ranges of the K, K_S, and K_N operators correspond to $\text{Span}\{f_m(t) - \bar{f}\}$, $\text{Span}\{f_m(t) - \bar{f}\}$, and $\text{Span}\{f_m(t) - \bar{f}_m\}$, respectively. We also know from Eqs. (25) and (26) that $\mathbf{S}$ and $\mathbf{N}$ are subspaces of $\mathbf{V}$, and therefore $\mathbf{S}$ and $\mathbf{N}$ have complementary subspaces (which could be empty) within $\mathbf{V}$, $\mathbf{S}^c$, and $\mathbf{N}^c$, where

$$\mathbf{S}^c \equiv \{\mathbf{v} \in \mathbf{V}; \; \mathbf{v}K_S\mathbf{v} = 0\} = \mathbf{V} \cap \text{Null}(K_S),$$
$$\mathbf{N}^c \equiv \{\mathbf{v} \in \mathbf{V}; \; \mathbf{v}K_N\mathbf{v} = 0\} = \mathbf{V} \cap \text{Null}(K_N). \quad (72)$$

Under Computation of the Eigenvectors and in part A of the Appendix, we argued that any canonical variate χ or generalized indicator ϕ can be uniquely determined up to the range of K, spanned by its eigenvectors with positive eigenvalues. Hence, χ and ϕ are members

of $\mathbf{V}$, and we consider $\mathbf{V}$ as our fundamental working space from which the signal images are sought.

Now, because of the identity (59), we have

$$\text{Rank}(K) = \text{Rank}(K_S + K_N). \quad (73)$$

Since $\mathbf{S} \subseteq \mathbf{V}$ and $\mathbf{N} \subseteq \mathbf{V}$, we can infer that the union of these subspaces cover $\mathbf{V}$ entirely (Fig. 8D),

$$\mathbf{V} = \mathbf{S} \cup \mathbf{N}. \quad (74)$$

Now, since the Q ratio is defined as $Q(\mathbf{v}) = \mathbf{v}K_S\mathbf{v}/\mathbf{v}K_N\mathbf{v}$, $Q(\mathbf{v}) = 0$ for any $\mathbf{v} \in \mathbf{S}^c$, $Q(\mathbf{v}) = \infty$ for any $\mathbf{v} \in \mathbf{N}^c$, and $0 < Q(\mathbf{v}) < \infty$ for $\mathbf{v} \in \mathbf{S} \cap \mathbf{N}$. Also from Eq. (73), it follows that

$$\text{Rank}(K) \le \text{Rank}(K_S) + \text{Rank}(K_N), \quad (75)$$

$$0 \le \text{Rank}(K) - \text{Rank}(K_N) \le \text{Rank}(K_S). \quad (76)$$

This inequality implies that $\mathbf{N}^c$ can have at most $\text{Rank}(K_S)$ dimensions, thus canonical variate χ can exist within $\mathbf{V}$ such that $\chi K_N \chi$ becomes 0, that is, $Q(\chi) \to \infty$.

For example, let us consider the simplest (yet perhaps most general) situation, in which the set of all data images $\{f_m(t)\}$ is linearly independent. This is indeed the most likely case in a typical optical imaging experiment, in which the total number of pixels per image, P, far exceeds the number of collected images, M, that is, $MT \ll P$. In the presence of random photon noise, etc., the set $\{f_m(t)\}$ tends to be linearly independent and forms MT dimensional subspace. In this scenario, it is straightforward to calculate the exact dimensionality of the subspaces, $\mathbf{V}$, $\mathbf{S}$, and $\mathbf{N}$, vis.,

$$\dim(\mathbf{V}) = \text{Rank}(K) = MT - 1,$$

$$\dim(\mathbf{S}) = \text{Rank}(K_S) = M - 1, \quad (77)$$

$$\dim(\mathbf{N}) = \text{Rank}(K_N) = M(T - 1),$$

where M is the number of conditions and T is the number of samples per condition. It is now evident from the relation (73) that K_N becomes "inevitably" more rank-deficient than K by precisely $M - 1$, and this creates a subspace $\mathbf{N}^c$ of dimension $M - 1$ in which the Q ratios become unbounded.

We should remark that this feature is independent of the data's signal content; an unbounded Q could occur even if the dataset carried absolutely no signal in it. However, an infinite Q could also occur due to the presence of a noiseless signal, as we witnessed under A Nonstochastic Signal. Therefore, when one obtains an infinite Q, one cannot conclude whether it represents a real signal or merely an artifact. One might call such dataset "ill-conditioned" but unfortunately, this is of-

ten the case in optical imaging experiments. This observation casts shadows on the validity of a signal identification strategy based solely on the signal-to-noise ratio.

The generalized indicator method, on the other hand, identifies signals by the weighted signal–noise difference (Eq. (12)) and relies on the Q ratio only as a surrogate statistic to score the purity of the extracted signal. In effect, the generalized indicator analysis seeks an image ϕ in $\mathbf{S}$ such that ϕ is as parallel to $\mathbf{S}$ as possible while keeping Q above some threshold level, q_c. Hence, it allows solutions to lie inside the space $\mathbf{S} \cap \mathbf{N}$, even when a nonempty $\mathbf{N}^c$ exists in $\mathbf{V}$. This feature becomes especially important when the true signal in question actually has some intrinsic fluctuations, as is the case with many biological systems. A vector $\mathbf{v}$ of a noisy signal, by definition, belongs to the intersection $\mathbf{S} \cap \mathbf{N}$, i.e., $0 < Q(\mathbf{v}) < \infty$. While the generalized indicator method is capable of capturing such a signal, the canonical analysis will completely miss the signal, as demonstrated under A Stochastic Signal, because it yields $Q(\chi) \to \infty$ by forcing χ to be orthogonal to $\mathbf{N}$.

ACKNOWLEDGMENTS

REFERENCES

Arieli, A., Shoham, D., Hildesheim, R., and Grinvald, A. 1995. Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *J. Neurophysiol.* **73:** 2072–2093.

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. 1996. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* **273:** 1868–1871.

Bell, A. J., and Sejnowski, T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7:** 1129–1159.

Bellman, R. 1960. *Introduction to Matrix Analysis.* McGraw–Hill, New York.

Blasdel, G. 1992. Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J. Neurosci.* **12:** 3115–3138.

Blasdel, G., and Salma, G. 1986. Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* **321:** 579–585.

Boch, R. D. 1975. *Multivariate Statistical Methods in Behavioral Research.* McGraw–Hill, New York.

Bonhoeffer, T., and Grinvald, A. 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* **353:** 429–431.

Box, G. E. P., Hunter, W., and Hunter, J. S. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* Wiley, New York.

Everson, R. E., Knight, B. W., and Sirovich, L. 1997. Separating spatially distributed response to stimulation from background. I. Optical imaging. *Biol. Cybernet.* **77:** 407–417.

Everson, R. E., Prashanth, A. K., Gabbay, M., Knight, B. W., Sirovich, L., and Kaplan, E. 1998. Representation of spatial frequency and orientation in the visual cortex. *Proc. Natl. Acad. Sci. USA* **95:** 8334–8338.

Fisher, R. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7:** 179–188.

Friston, K. J., Frith, C. D., Frackowiak, R. S. J., and Turner, R. 1995. Characterizing dynamic brain responses with fMRI: A multivariate approach. *NeuroImage* **2:** 157–151.

Frostig, R. D., Lieke, E. E., Ts'o, D. Y., and Grinvald, A. 1990. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by *in vivo* high resolution optical imaging of intrinsic signals. *Proc. Natl. Acad. Sci. USA* **87:** 6082–6086.

Gabbay, M., Brennan, C., Kaplan, E., and Sirovich, L. 2000. A principal components-based method for the detection of neuronal activity maps: Application to optical imaging. *NeuroImage* **11:** 313–325.

Gegiu, L. M., Orbach, D., Yokoo, T., Sornboger, A., Sirovich, L., and Kaplan, E. 2001. Color discrimination in area V1 of the macaque visual cortex: An optical imaging study. Manuscript in preparation.

Grinvald, A. 1992. Optical imaging of architecture and function in the living brain sheds new light on cortical mechanisms underlying visual perception. *Brain Topogr. Winder.* **5:** 71–75.

Hyvarinen, A., and Oja, E. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* **9:** 1483–1492.

Issa, N. P., Trepel, C., and Stryker, M. P. 2000. Spatial frequency maps in cat visual cortex. *J. Neurosci.* **20:** 8504–8514.

Malonek, D., and Grinvald, A. 1996. The spatial and temporal relationship between cortical electrical activity and responses of the microcirculation during sensory stimulation: Implications for optical, PET, and MR functional brain imaging. *Science* **272:** 551–554.

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I., and Grinvald, A. 1997. Vascular imprints of neuronal activity: Relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. *Proc. Natl. Acad. Sci. USA* **94:** 14826–14831.

Orbach, H. S., and Cohen, L. B. 1983. Optical monitoring of activity from many areas of the in vitro and in vivo salamander olfactory bulb: A new method for studying functional organization in the vertebrate central nervous system. *J. Neurosci.* **3:** 2251–2262.

Sailstad, C. 2001. Functional architecture of the cat visual cortex: Role of the reciprocal pathway from area 18. Manuscript in preparation.

Schiessl, I., Stetter, M., Mayhew, J. E., McLoughlin, N., Lund, J. S., and Obermayer, K. 2000. Blind signal separation from optical imaging recordings with extended spatial decorrelation. *IEEE Trans. Biomed. Eng.* **47:** 573–577.

Searle, S., Casella, R. G., and McCulloch, C. 1992. *Variance Components.* Wiley, New York.

Sirovich, L. 1987. Turbulence and the dynamics of coherent structures. I. Coherent structures. II. Symmetries and transformations. III. Dynamics and scaling. *Q. Appl. Math.* **XLV:** 561–590.

Sirovich, L., and Everson, R. E. 1992. Analysis and management of large scientific databases. *Int. J. Supercomput. Appl.* **6:** 50–68.

Sirovich, L., Everson, R. E., Kaplan, E., Knight, B. W., O'Brien, E., and Orbach, D. 1996. Modeling the functional organization of the visual cortex. *Physica D* **96** 355–366.

Sirovich, L., and Kaplan, E. 2001. Analysis methods for optical imaging. In *Methods for in Vivo Optical Imaging of the Central Nervous System* (R. Frostig, Ed.). CRC Press, Boca Raton, FL, in press.

Shmuel, A., and Grinvald, A. 1996. Functional organization of mo-

tion and its relationship to orientation maps in cat area 18. *J. Neurosci.* **16:** 6945–6964.

Stetter, M., Schhiessl, I., Otto, T., Sengpiel, F., Hbener, M., Bonhoeffer, T., and Obermayer, K. 2000. Principal component analysis and blind separation of sources for optical imaging of intrinsic signals. *NeuroImage* **1:** 482–490.

Ts'o, D. Y., Frostig, R. D., Lieke, E. E., and Grinvald, A. 1990. Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science* **249:** 417–420.

Weliky, M., Bosking, W., and Fitzpatrick, D. 1996. A systematic map of direction preference in primary visual cortex. *Nature* **379:** 725–728.