A Unifying View of Second-Order Measures of Quality for Signal Classification

WILLIAM A. GARDNER, MEMBER, IEEE

***Abstract*—The major purpose of this paper is to promote interchange between the fields of pattern recognition and communications, in the realm of statistical classification. The general class of *second-order measures of quality* for statistical classification is defined. The variety of members in this class that have been used by practitioners or proposed by theorists for numerical pattern-classification and signal waveform-classification are compared and contrasted. The several measures that are the most generally applicable are shown to be either equivalent to each other or characterizable in terms of each other, thereby revealing an inherent unity. For example, the ratio of between-class-scatter to within-class-scatter used in pattern recognition and the ratio of signal-energy to noise-energy used in communications are unified through an identification of *signal* with between-class-scatter and *noise* with within-class-scatter. Results on equivalences are stated and proved for waveform classification rather than numerical classification in order to complement the extensive literature on the latter, and to emphasize applicability to communications. This entails introduction of a scatter ratio for waveforms. In a companion paper,[1] second-order measures of quality are used as a basis for a *general nearest-prototype signal-classification* methodology; canonical signal features for this methodology are identified, and a general approach for determining appropriate class prototypes is given. These two papers provide an integrated approach to the design of a complete signal classifier, i.e., feature extraction and discriminant-functional design tailored to fit a minimum-distance discrimination rule.**

I. BACKGROUND AND PURPOSE

FOR many problems of statistical classification, the Bayes risk (e.g., probability of misclassification) is the most desirable measure of performance of a classifier, and minimum Bayes risk is the most desirable criterion for classifier design. However, in many practical applications, the problem of obtaining an explicit solution for the minimum-risk classifier is intractable, or is impossible because of incomplete specification of the underlying probabilistic model. Similarly, the problem of evaluating the risk associated with a given classifier can be difficult or impossible. As a result, various alternative design criteria and performance measures have been proposed and used in practice. Perhaps the most popular family of alternatives that are not entirely ad hoc, and that have been studied and used by theorists and practitioners is the family that is based on *second-order measures of quality*. A second-order measure of quality is a measure that is defined completely in terms of only second-order probabilistic parameters, viz., means and covariances.

In the field of communications, mean-squared error (MSE), and various measures of receiver output signal-to-noise ratio (SNR) are second-order measures of quality that are commonly used for design of signal classifiers, e.g., detectors. Although a general approach to signal classifier design based on MSE has recently been proposed [20], no such general approach based on SNR has yet been proposed, although both MSE and SNR have been used for specific design problems (cf. [2], [19], [23], [29, Section 7.3], [32], [39].) In the field of pattern recognition, MSE and various measures of the ratio of between-class scatter to within-class scatter are second-order measures of quality that are commonly used for design of pattern discriminants (cf. [8], [10], [11, Sections 4.10-4.12, 5.8, 5.12, 5.13], [17, Section 4.3], and references therein), extraction of pattern features (cf. [8], [10], [17, Section 9.2], [18], and references therein), and design of clustering algorithms (cf. [11, Section 6.8], [17, Section 11.2], and references therein). A specific objective of this paper is to show that there is a general approach to classifier design that is based on SNR and that is equivalent to one form of the general approach proposed in [20] based on MSE; and, furthermore, that this SNR approach is a natural extension of a well-known general approach in pattern recognition, which is based on a scatter ratio.

As design criteria, second-order measures have been successfully applied when the set of discriminant functionals (constraint space), within which an optimum is sought, contains only linear functionals or their generalizations which are defined in terms of prescribed linear spaces that are either finite dimensional ([11, Section 5.3], [33, Section 2.11], [42, Sections 3.3 and 3.4]) or infinite dimensional [20]. Although these second-order measures have also been used to predict the performance of discriminant functionals and feature sets for pattern and signal classification (cf. [8], [10], [11], [13], [17], [32], [39], [45], [47, Section 4.3], and references therein), they have met with more limited success as performance predictors (cf. [24]). As a result, a premise of this paper is that second-order measures are of practical value for deriving *candidate* classifiers, but since second-order measures can be poor predictors of Bayes risk, the acceptance, rejection, or modification of a candidate classifier should, whenever feasible, be based on an evaluation of the risk associated with the candidate. Nevertheless, in applications where risk cannot be evaluated or approximated, second-order measures might have to be employed to assess classifier quality (e.g., via risk bounding [8], [10]).

Motivated partly by the analytical tractability and partly by

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received March 17, 1979; revised January 5, 1980.

The author is with the Signal and Image Processing Laboratory, Department of Electrical and Computer Engineering, University of California, Davis, CA 95616.

[1] The companion paper, "Nearest-prototype signal classification," which will appear in the IEEE TRANSACTIONS ON INFORMATION THEORY, subsumes [17] in [20], which was in preparation at the time of publication [20].

the practical utility of second-order measures of quality, a number of investigators have contributed to the development of theoretical justifications for their use when the Bayes risk measure would be the most desirable but cannot be applied. Although there has been some work on justification in the communications literature (cf. [20], [36], [39], [41]), the majority has appeared in the pattern recognition literature (cf. [8], [10], [11], [17], [18], [28], [34], [42], [50]). In addition to the contributions on justifying individual second-order measures, some results on unification of various second-order measures have been reported in the pattern recognition literature. Results include equivalences among various design criteria (cf. [8], [10], [11, Section 5.8], [17, Sections 9.2, and 11.2], [18], [42, Section 3.2]), convergence of *probabilistic* second-order-optimal discriminant functionals to Bayes optimal discriminant functionals as the constraint space of generalized linear functionals expands to include (in the limit) all potentially optimal functionals (cf. [18], [42, Sections 3.2 and 3.3]), and convergence of *empirical* second-order-optimal (e.g., least squares and stochastic approximation) discriminant functionals to probabilistic second-order-optimal (e.g., minimum MSE) discriminant functionals as the number of samples increases without limit (cf. [11, Section 5.8], [28], [34], [50]).

It should be clarified at this point that an empirical second-order measure of quality is defined in terms of empirical averages, viz., sample-mean and sample-covariance statistics (e.g., Fisher's discrimination measure [11, Section 4.10], [15]), whereas the probabilistic counterpart is defined in terms of parameters of probability distributions, viz., mean and covariance. The empirical measures have been more common in pattern recognition applications, and the probabilistic measures have been more common in communications applications with the exception of the rapidly developing adaptive signal classification area.

Although there is a substantial and growing interchange of statistical and probabilistic concepts and methods between the fields of pattern recognition and communications (cf. [5, ch. 12 and references therein]), the theory of second-order measures of quality for classification has been developed mostly within the field of pattern recognition, and this has left a gap between these two fields. A unified treatment of the theory of second-order measures of quality that emphasizes signal classification and detection would narrow this gap and promote interchange. This emphasis requires extension and generalization of the established theory for a finite number of observables (numerical classification) to a theory for a nondenumerable infinity of observables (waveform classification).

With this brief review as background, the main purpose of this paper can now be stated. In addition to the preceding specific objective, the general objective is to extend, generalize, and unify the theory of second-order measures of quality for signal classification. Specifically, the variety of second-order measures of quality that have been used or proposed for numerical and waveform classification are defined in the Appendix, and are compared and contrasted there and in Sections II and III. The several measures that are the most generally applicable are shown in Section III to be either equivalent to each other or characterizable in terms of each other, thereby revealing an inherent unity. In a companion paper,[1] a general *nearest-prototype signal-classification* methodology that is based on second-order measures of quality is proposed; and canonical signal features, and appropriate signal discriminants for this methodology are determined. These two papers provide an integrated approach to the design of a complete signal classifier, i.e., feature extraction and discriminant-functional design taylored to fit a minimum-distance signal discrimination rule.

Although only fixed-sample classification is treated, second-order measures of quality are useful for sequential classification as well. For example, approaches to linear least squares sequential classification are described in [43], [44].

II. SECOND-ORDER MEASURES OF QUALITY

In Section II-A, the waveform classification problem is formulated, and the generalized linear discriminant functional is defined. Then, in Section II-B, the two major second-order measures of quality for binary signal classification are defined. These two measures are shown in the Appendix to be superior or equivalent to, or to include as special cases, the majority of the 12 second-order measures of quality that have been proposed and/or used in practice. It is shown in Section III that these two major measures of quality yield equivalent discriminant functional design criteria. In Section II-C, these binary (two class) discriminant functional design criteria are generalized to M-ary (multiclass) design criteria. Then a single-valued alternative to these M-valued measures of quality is defined. It is shown in Section III that this single-valued measure can be characterized by each of the two M-valued measures.

A. Generalized Linear Discriminant Functionals

We consider observations consisting of a waveform $y = \{y(t): t \in T\}$ on a time-interval T. The waveform y is a sample path from a continuous parameter random process Y. The probability measure μ for Y is a mixture of M measures $\{\mu_i\}_1^M$

$$\mu = \sum_{i=1}^{M} p_i \mu_i. \tag{1}$$

We are concerned with an M-ary communication system for which we have the following probabilistic model: the ith of M available signals is selected with probability p_i for transmission, and the corresponding received waveform is a sample from the ith class C_i which is defined to be the class of all sample paths from the random process $Y \mid C_i$, which has probability measure μ_i. Thus, the problem of deciding which of the M signals was transmitted is a Bayesian classification problem with prior class-probabilities $\{P[C_i]\}_1^M = \{p_i\}_1^M$, and posterior class-probabilities denoted by $\{P[C_i \mid y]\}_1^M$.

A discriminant functional Φ is a functional that maps the observed waveform y into an N-tuple x of real numbers $\{x_j\}_1^N$ that can be used to discriminate amongst the M possible classes, i.e., to classify y:

$$x = \Phi(y) \in R^N$$
$$x_j = \Phi_j(y) \in R \qquad j = 1, 2, \cdots, N. \tag{2}$$

The N-tuple x is referred to as a *discriminant.* If the functional is constrained to be continuous and linear, the Riesz representation theorem [12] can be invoked to represent it in inner product form

$$x_j = (y, \phi_j) \triangleq \int_T \phi_j(t) y(t)\, dt \quad j = 1, 2, \cdots, N. \tag{3}$$

More generally, if the functional is constrained to be the composition of a continuous linear functional with a prescribed nonlinear transformation g, it is called a *generalized linear discriminant functional*, and it can be represented by

$$x_j = (z, \phi_j)_\Lambda \qquad j = 1, 2, \cdots, N \tag{4}$$
$$z = g(y),$$

where $(\cdot, \cdot)_\Lambda$ denotes inner product on the linear space Λ containing the images of the observed waveforms y (sample paths of Y) under the transformation g.

An example of a generalized linear discriminant functional is a generalized polynomial functional. As a specific example, the transformation g corresponding to a second degree generalized polynomial is

$$g(\{y(t): t \in T\}) = \{1, y(t), y(t_1) y(t_2): t, t_1, t_2 \in T\},$$

and (4) becomes (with $\phi_j = \{\phi_j^0, \phi_j^1, \phi_j^2\}$)

$$x_j = \phi_j^0 + \int_T \phi_j^1(t) y(t)\, dt + \int_T \int_T \phi_j^2(t_1, t_2) y(t_1) y(t_2)\, dt_1\, dt_2. \tag{5}$$

Another example of a generalized linear discriminant functional is a parallel connection of nonlinearities followed by correlators, i.e.,

$$x_j = \sum_{i=1}^{M} \int_T g_i[y(t_i), t_i]\, \phi_j^i(t_i)\, dt_i, \tag{6}$$

for which the transformation g is

$$g(\{y(t): t \in T\}) = \{\{g_i[y(t_i), t_i] : t_i \in T\}\}_1^M.$$

Specific examples of nonlinearities g_i are *clippers, limiters,* and

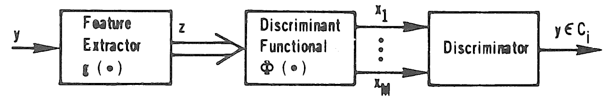


Fig. 1. Signal classifier.

hole-punchers with possibly time-varying biases $b_i(t)$, e.g.,

$$g_i[y(t), t] = \begin{cases} a_i & \text{for } y(t) - b_i(t) > 0 \\ -a_i & \text{for } y(t) - b_i(t) \leqslant 0. \end{cases}$$

Discriminant functionals (3), (5), and (6) include signal detector structures of demonstrated practical value. In the pattern recognition literature g is often interpreted as a feature extractor when its range is contained in finite dimensional Euclidean space; this interpretation is used in Section III, and expanded on in a companion paper[1] (see Fig. 1).

A second-order measure of quality for a random discriminant X is defined in terms of the following second-order probabilistic parameters.[2]

1) The conditional mean N-tuple $m_i(X)$ with (j)th element

$$m_i(j) \triangleq E\{X_j \mid C_i\} = \int \Phi_j(Y)\, d\mu_i. \tag{7}$$

2) The conditional $N \times N$ covariance matrix $k_i(X)$ with (jq)th element

$$k_i(j, q) \triangleq E\,\{[X_j - m_i(j)]\,[X_q - m_i(q)] \mid C_i\}$$
$$= \int [\Phi_j(Y) - m_i(j)]\,[\Phi_q(Y) - m_i(q)]\, d\mu_i. \tag{8}$$

The notation of $m_i(X)$ and $k_i(X)$ includes X explicitly so that these quantities can be distinguished from the $L^2[T]$ vector $m_i(Y)$ and $L^2[T]$ operator $k_i(Y)$

$$m_i(t) \triangleq E\{Y(t) \mid C_i\} \tag{9}$$
$$k_i(t, \tau) \triangleq E\{[Y(t) - m_i(t)]\,[Y(\tau) - m_i(\tau)] \mid C_i\}. \tag{10}$$

For a linear discriminant functional, the N-tuple $m_i(X)$ and $N \times N$ matrix $k_i(X)$ can be expressed explicitly in terms of the discriminant functional Φ and the $L^2(T)$ vector $m_i(Y)$ and $L^2(T)$ operator $k_i(Y)$ as follows

$$m_i(j) = \int_T \phi_j(t) m_i(t)\, dt \tag{11}$$

$$k_i(j, q) = \int_T \int_T \phi_j(t) \phi_q(\tau) k_i(t, \tau)\, dt\, d\tau, \tag{12}$$

[2] Some second-order measures depend on the *joint* second-order probabilistic parameters of X *and* a prescribed random variable [e.g., S in (A-3)]. But as shown in a companion paper,[1] the second-order optimal discriminant for such a measure can still be completely specified in terms of only (7) and (8) when S is appropriately prescribed.

or alternatively in $L^2(T)$ inner product notation

$$m_i(j) = (m_i(Y), \phi_j) \tag{13}$$

$$k_i(j, q) = (k_i(Y) \cdot \phi_q, \phi_j). \tag{14}$$

Similarly, representations (13) and (14) hold for generalized linear discriminant functionals if Y is replaced by Z and $(\cdot, \cdot)$ is replaced by $(\cdot, \cdot)_\Lambda$. Thus, a second-order measure of quality for a linear discriminant functional is defined in terms of the discriminant functional representors $\{\phi_i(t)\}_1^N$ and the mean functions $\{m_i(t)\}_1^M$ and covariance functions $\{k_i(t, \tau)\}_1^M$, and similarly for generalized linear discriminant functionals. (The specific examples of first and second degree generalized polynomials are considered in detail in [20].) The general representations (13) and (14) for the second-order probabilistic parameters of the random discriminant X, in terms of the discriminant functions, $\{\phi_i\}$ and the mean and covariance functions of the observed waveform Y (or its transformed version Z), simplify the general analysis of second-order measures of quality as illustrated in Sections II and III.

The preceeding discussion pertains to probabilistic second-order measures of quality. Empirical measures are defined in terms of second-order empirical parameters obtained by replacing the expected values in (7) and (8) with the following empirical averages:

$$\tilde{m}_i(j) \triangleq \frac{1}{N_i} \sum_{n=1}^{N_i} x_j^i(n) \tag{15}$$

$$\tilde{k}_i(j, q) \triangleq \frac{1}{N_i} \sum_{n=1}^{N_i} [x_j^i(n) - \tilde{m}_i(j)]\,[x_q^i(n) - \tilde{m}_i(q)] \tag{16}$$

where $\{x_j^i(n)\colon n = 1, 2, \cdots, N_i\}$ is a set of samples of the class-conditional random variable $X_j \mid C_i$. Similarly, the prior probabilities $\{p_i\}_1^M$ that are sometimes used in probabilistic measures of quality are replaced with the sample estimates

$$\tilde{p}_i \triangleq N_i \left[\sum_{j=1}^{M} N_j\right]^{-1} \tag{17}$$

where N_i is the number of samples drawn from the total population (μ) that belong to class C_i (μ_i).

B. Binary Classification

Since Bayes-optimal classification can be accomplished with a discriminant functional Φ having only $N = M$-1 output discriminants (viz., any M-1 of the M posterior class probabilities), second-order measures of quality for binary $(M = 2)$ classification are most often defined for a discriminant functional with a scalar $(N = 1)$ output X. From the 12 second-order measures of quality discussed in the Appendix, the following three emerge as worthy of further consideration for general use. (In these definitions, the standard notation $m_i(X)$ and $\sigma_i^2(X)$ is used in place of $m_i(1)$ and $k_i(1, 1)$, from (7) and (8), for this scalar case.)

1) Generalized signal-to-noise ratio (GSNR):

$$\text{GSNR}\,(X) = \frac{[m_1(X) - m_2(X)]^2}{w_1 \sigma_1^2(X) + w_2 \sigma_2^2(X)} \tag{18}$$

where $\{w_1, w_2\}$ is a nonnegative weight vector. It can be shown[3] that when $w_i = p_i$,

$$\text{GSNR} = \text{SNR}/(1 + p_1 p_2\, \text{SNR}) \tag{19}$$

where

$$\text{SNR} \triangleq \frac{[m_1(X) - m_2(X)]^2}{\sigma^2(X)}, \tag{20}$$

and $\sigma^2(X)$ is the unconditional variance of X. As a result of (19), the design criteria of maximization of GSNR and SNR (w.r.t. Φ) are equivalent. We shall henceforth focus on SNR.

2) Mean-squared error (MSE):

$$\text{MSE}\,(X) \triangleq E\{(S - X)^2\} \tag{21}$$

where S is a random signal parameter with unique (nonrandom) class-conditional values (see (22) for an example), and X is an estimator for S (denoted by $X = \hat{S}$). As discussed in a companion paper,[1] the appropriateness of MSE depends on the prescription of the signal parameter S for M-ary $(M > 2)$ classification but is invariant (as a design criterion) to S for binary classification. The most commonly used prescription (see [20]) is the class indicator[4]

$$S = \delta(C_1) \triangleq \begin{cases} 1 & \text{for } y \in C_1 \\ 0 & \text{for } y \notin C_1. \end{cases} \tag{22}$$

It can be shown[3] that the design criterion of minimization (w.r.t. Φ) of MSE (X) with $S = \delta(C_1)$ is equivalent to the same criterion but with $S = P[C_1 \mid Y]$. This is particularly interesting since the "signal parameter" $S = P[C_1 \mid Y]$ is a functional of the observations and does not have unique class-conditional values. Other prescriptions for S are binary-valued signal modulation parameters such as amplitude for an ASK signal, or frequency for an FSK signal [19], [20], [23].

C. M-ary Classification

The two second-order measures of quality SNR and MSE for binary classification can be applied to M-ary classification simply by interpreting M-ary classification in terms of M binary classifications (cf. [11, Section 5.12.3]). For this purpose, we consider a discriminant functional Φ with $N = M$ (rather than M-1) component discriminants $\{X_i\}_1^M$, and we define SNR and MSE for each as follows.

[3] Concise proofs are given in the unpublished report, "On minimum-MSE and maximum-SNR signal discriminants," which is available upon request from the author.

[4] Our notational convention is violated here since we use lower case δ, rather than capital Δ, to denote a random variable.

$$\mathrm{SNR}_i\,(X_i) \triangleq \frac{[m_i(X_i) - m_i^c(X_i)]^2}{\sigma^2(X_i)} \tag{23}$$

$$\mathrm{MSE}_i\,(X_i) \triangleq E\{(S_i - X_i)^2\} \tag{24}$$

where, for example, $S_i = \delta(C_i)$ or $P[C_i \mid Y]$. In (23), $m_i^c(X_i)$ is the "complement" of $m_i(X_i)$, i.e., the mean of X_i conditioned on *not* C_i (which is denoted by C_i^c). In (24), the expectation is unconditional. As an alternative to such M-valued second-order measures of quality, a single-valued measure of scatter ratio can be used. As discussed in the Appendix, of the six most well-known measures of scatter ratio, the following turns out to be the only one that is closely related to SNR and MSE for the M-ary ($M > 2$) classification problem:

$$J(X) \triangleq \mathrm{trace}\,\{S_T^{-1} S_B\} \tag{25}$$

where S_B is the between-class scatter matrix

$$S_B \triangleq \sum_{i=1}^{M} w_i p_i [m_i(X) - m(X)]\,[m_i(X) - m(X)]^t \tag{26}$$

with weight vector $\{w_i\}$ ($m(X)$ is the unconditional mean vector), and S_T is the total (within-class plus between-class) scatter matrix with unit weight vector

$$S_T = \sum_{i=1}^{M} p_i k_i(X) + \sum_{i=1}^{M} p_i [m_i(X) - m(X)] \cdot [m_i(X) - m(X)]^t \tag{27}$$

where $m_i(X)$ and $k_i(X)$ are the N-tuple and $N \times N$ matrix with elements defined by (7) and (8), respectively. The symbol $[\cdot]^t$ denotes matrix transposition. It is easily shown that J can be reexpressed by

$$J(X) \triangleq \sum_{i=1}^{M} p_i w_i [m_i(X) - m(X)]^t \cdot k(X)^{-1} [m_i(X) - m(X)] \tag{28}$$

where $k(X)$ is the unconditional covariance matrix.

In order to relate J to SNR, we extend the definition (23) from the scalar case X to the N-tuple case X

$$\mathrm{SNR}_i\,(X) \triangleq [m_i(X) - m_i^c(X)]^t \cdot k(X)^{-1} [m_i(X) - m_i^c(X)], \tag{29}$$

and we extend the definitions (28) and (29) from the N-tuple case (numerical measures) to the general vector case (waveform measures) by using general inner product notation

$$J(Z) \triangleq \sum_{i=1}^{M} p_i w_i (m_i(Z) - m(Z),\quad k(Z)^{-1} \cdot \{m_i(Z) - m(Z)\})_\Lambda \tag{28$'$}$$

$$\mathrm{SNR}_i\,(Z) \triangleq (m_i(Z) - m_i^c(Z), k(Z)^{-1} \cdot \{m_i(Z) - m_i^c(Z)\})_\Lambda. \tag{29$'$}$$

Similarly, these measures can be applied to the untransformed observations Y simply by replacing Z with Y and letting the inner product $(\cdot, \cdot)_\Lambda$ be the $L^2(T)$ inner product. Now, using the identity $m_i - m = p_i^c(m_i - m_i^c)$, we obtain

$$J(Z) = \sum_{i=1}^{M} w_i p_i (p_i^c)^2\ \mathrm{SNR}_i\,(Z) \tag{30}$$

where $p_i^c = 1 - p_i$. Thus, if we prescribe the weight vector $w_i = p_j \delta_{ij}$ (where δ_{ij} is the Kronecker delta), then

$$J(Z) = (p_j p_j^c)^2\ \mathrm{SNR}_j\,(Z). \tag{31}$$

These relationships between the *waveform scatter ratio* J and SNR are used in Section III to relate $J(X)$ to $\{\mathrm{SNR}_i(X_i)\}_1^M$ and $\{\mathrm{MSE}_i(X_i)\}_1^M$.

III. EQUIVALENCES

In this section, it is shown that MSE and SNR yield equivalent discriminant functional design criteria, and that J, min-MSE and max-SNR yield equivalent feature extractor design criteria. It is also shown that all these design criteria are consistent in the sense that they yield Bayes-optimal designs when no constraints are imposed on the feature extractor $g(\cdot)$ or the discriminant functional $\Phi(\cdot)$.

It can be shown [3] that the generalized linear discriminant that minimizes MSE for the prescription $S_i = \delta(C_i)$ in (24) is

$$\Phi_{\mathrm{MSE}_i}(y) = \hat{\delta}(C_i) = \hat{P}[C_i \mid y] = p_i\{1 + ([z - m(Z)],\ k(Z)^{-1} \cdot [m_i(Z) - m(Z)])_\Lambda\}, \tag{32}$$

and that which maximizes SNR is

$$\Phi_{\mathrm{SNR}_i}(y) = \alpha_i + \beta_i([z - m(Z)], k(Z)^{-1} \cdot [m_i(Z) - m_i^c(Z)])_\Lambda \tag{33}$$

where $z = g(y)$, and α_i and β_i are arbitrary scalars. It follows from (32) and (33) that Φ_{SNR_i} and Φ_{MSE_i} can be related by

$$\Phi_{\mathrm{SNR}_i}(y) = \alpha_i + \beta_i(\Phi_{\mathrm{MSE}_i}(y)/p_i - \Phi_{\mathrm{MSE}_i}^c(y)/p_i^c). \tag{34}$$

If the transformed observations Z are centered ($m(Z) = 0$), then $m_i^c(Z) = -p_i(p_i^c)^{-1} m_i(Z)$, and (34) reduces to

$$\Phi_{\mathrm{SNR}_i}(y) = \alpha_i + \beta_i(p_i p_i^c)^{-1}(\Phi_{\mathrm{MSE}_i}(y) - p_i). \tag{35}$$

Thus, except for scale factor and bias, Φ_{MSE_i} and Φ_{SNR_i} are the same; with the prescription $\alpha_i = p_i$ and $\beta_i = p_i p_i^c$, they are identical

$$\Phi_{\mathrm{SNR}_i}(y) = \Phi_{\mathrm{MSE}_i}(y). \tag{36}$$

It can be shown[3] that if Φ is not constrained to be a generalized linear functional, but is allowed to be any functional with finite mean square images ($E\{\Phi(Y)\} < \infty$), then maximization of SNR yields the design

$$\Phi_{\mathrm{SNR}_i}(y) = \alpha_i + \beta_i\{P[C_i \mid y]/p_i - P[C_i^c \mid y]/p_i^c\}, \qquad (37)$$

and thus, $\{\Phi_{\mathrm{SNR}_i}(y)\}_1^M$ comprises a Bayes-optimal set of discriminants. Parallel to (37), it follows from (34) that the constrained max-SNR design (33) can be reexpressed as

$$\Phi_{\mathrm{SNR}_i}(y) = \alpha_i + \beta_i\{\hat{P}[C_i \mid y]/p_i - \hat{P}[C_i^c \mid y]/p_i^c\}. \qquad (37)'$$

It can be shown[3] that the maximum value of SNR_i resulting from the second-order-optimal design Φ_{SNR_i} is

$$\text{max-SNR}_i\,(X_i) = \mathrm{SNR}_i(Z) \qquad (38)$$

where $\mathrm{SNR}_i(Z)$ is defined by (29)′. Thus, it follows from (30) that the waveform scatter ratio J can be characterized by the maximum values of $\{\mathrm{SNR}_i(X_i)\}$

$$J(Z) = \sum_{i=1}^{M} w_i p_i (p_i^c)^2 \;\text{max-SNR}_i(X_i). \qquad (39)$$

Furthermore, it can be shown[3] that the minimum value of MSE_i resulting from the second-order-optimal design $\Phi_{\mathrm{MSE}_i}(y) = \hat{\delta}(C_i)$ is

$$\text{min-MSE}_i\,(\hat{\delta}(C_i)) = p_i p_i^c\,[1 - p_i p_i^c \;\text{max-SNR}_i\,(X_i)]. \qquad (40)$$

Hence, (39) and (40) yield the alternative characterization

$$J(Z) = \sum_{i=1}^{M} [w_i p_i - w_i p_i^{-1}\;\text{min-MSE}_i\,(\hat{\delta}(C_i))]. \qquad (41)$$

Furthermore, it can be shown[3] that

$$\sum_{i=1}^{M} \text{min-MSE}_i\,(\hat{\delta}(C_i)) = 1 - \|P\|_{ms}^2 + \sum_{i=1}^{M} \text{min-MSE}_i\,(\hat{P}[C_i \mid Y]) \qquad (42)$$

and, therefore (41), with $w_i = p_i$, becomes

$$J(Z) = \|P\|_{ms}^2 - \|p\|^2 - \sum_{i=1}^{M} \text{min-MSE}_i\,(\hat{P}[C_i \mid Y]) \qquad (43)$$

where

$$\|P\|_{ms}^2 \triangleq \sum_{i=1}^{M} E\{P[C_i \mid Y]^2\} \qquad (44)$$

$$\|p\|^2 = \sum_{i=1}^{M} p_i^2.$$

Finally, it follows directly from the preceeding that

$$J[g(Y)] \equiv J(Z) = J(X_*) \qquad (45)$$

where X_* is the second-order-optimal M-tuple of discriminants that maximize $\{\mathrm{SNR}_i(X_i)\}_1^M$ and equivalently minimize $\{\mathrm{MSE}_i(X_i)\}_1^M$ (with respect to $\{\phi_i\}$). Thus, it follows that as a feature extraction criterion, maximization of $J[g(Y)]$ with respect to $g(\cdot)$ is equivalent to maximization of $\{\text{max-SNR}_i(X_i)\}_1^M$ and to minimization of $\{\text{min-MSE}_i(X_i)\}_1^M$ for both $X_i = \hat{\delta}(C_i)$ and $X_i = \hat{P}[C_i \mid Y]$.

If $g(\cdot)$ is not constrained to some prescribed design set, then it follows from (43) and (45) that $J[g(Y)] = J(X_*)$ is maximized if and only if $X_i = \hat{P}[C_i \mid Y] = P[C_i \mid Y]$ (in mean square). Thus, maximum-J is a consistent design criterion for feature extraction. The maximum value of J is from (43)

$$\text{max-}J(Z) = \|P\|_{ms}^2 - \|p\|^2. \qquad (46)$$

A result similar to (36), but for the special case of only a finite number of observable random variables $\{Y_i\}$, was obtained by Sebestyen [42, Sections 3.2, 3.3]. Also, since Fisher's measure of quality is the empirical version of GSNR, and the sum-of-squared-errors is the empirical version of the MSE measure of quality, it is not surprising, in view of (36) and the equivalence of the maximum-SNR and maximum-GSNR criteria, that Fisher's criterion is equivalent to the least-squares criterion (cf. [11, Section 5.8]). Furthermore, an application of the law of large numbers yields asymptotic equivalence between these probabilistic and empirical counterparts (cf. [11, Section 5.8]). Results similar to (37), but for the special case of only a finite number of observables for which a probability density exists, have been obtained by R. Hines (cf. Sebestyen (1960) and Section 3.2, both in [42]) and P. Rudnick [41]. Results similar to (43)–(46), but for the special case of only a finite number of observables for which a probability density exists, have been obtained by Devijver [8], [10], and Fukunaga and Ando [18].

In conclusion, in addition to introducing measures of quality J and SNR, (28)′ and (29)′, that apply to waveform observations and features as well as numerical observations and features, we have generalized previously obtained equivalences,[5] and extended them in a way that unifies the various equivalences and the various approaches to discriminant functional

[5] The bulk of results on equivalence in Section III were obtained and submitted for publication in another journal prior to publication of [18].

design and feature extraction as they have been applied to both signal detection receiver design and pattern recognition machine design (cf. Appendix). In a companion paper,[1] we go one step further by proposing a general approach for designing a discriminator that uses the second-order-optimal discriminants discussed in this section to carry out the final stage of classification.

IV. CONCLUSIONS

Some of the literature cited in the Appendix indicates that some investigators do not take advantage of the direct links between their work, which is based on a specific second-order measure of quality, and the large body or results (cited in Section I) based on related, equivalent, or more general second-order measures of quality. It is felt that this is due in large part to insufficient cross-referencing and differences in terminology between the fields of pattern recognition and communications. This gap between fields in unfortunate because the general suboptimality of second-order measures of quality can yield misleading results in applications to both classifier design and performance evaluation. Thus, it is hoped that the unifying view of second-order measures of quality presented in this paper will promote interchange and help narrow the gap.

APPENDIX
CONVENTIONAL SECOND-ORDER MEASURES OF QUALITY

A. Binary Classification

The following is a comprehensive list of second-order measures of quality together with references to their application and in some cases, references to investigations of the measures themselves. The notation in Section II-B is used here.

1) Elementary SNR:

$$\mathrm{ESNR} = \frac{[m(X)]^2}{\sigma^2(X)}. \qquad \text{(A-1)}$$

For detection of a sure signal s in additive zero mean noise N using a linear discriminant functional, $X \mid C_1 = \Phi(s) + \Phi(N) \triangleq X_s + X_N$ and $X \mid C_2 = \Phi(N) \triangleq X_N$, and ([46, Section 4.2])

$$\mathrm{ESNR} = \frac{[X_s]^2}{\sigma^2(X_N)}. \qquad \text{(A-1)}'$$

2) Mean Square Ratio:

$$\mathrm{MSR} \triangleq \frac{m_1^2(X) + \sigma_1^2(X)}{m_2^2(X) + \sigma_2^2(X)}. \qquad \text{(A-2)}$$

For the classification problem described in 1)

$$\mathrm{MSR} = 1 + \frac{[X_s]^2}{\sigma^2(X_N)}, \qquad \text{(A-2)}'$$

and for the problem of detection of a random signal S in additive independent random noise N using a linear discriminant functional, $X \mid C_1 = X_S + X_N$ and $X \mid C_2 = X_N$, and

$$\mathrm{MSR} = 1 + \frac{E\{X_S^2\}}{E\{X_N^2\}}. \qquad \text{(A-2)}''$$

3) Mean-Squared-Error SNR:

$$\text{MSE-SNR} \triangleq \frac{E\{S^2\}}{E\{(S-X)^2\}} \qquad \text{(A-3)}$$

where X is interpreted as an estimate of a random signal parameter S([1], [4], [19], [20], [23], [25], [29, Section 7.3]).

4) Deflection:

$$D \triangleq \frac{[m_1(X) - m_2(X)]^2}{\sigma_2^2(X)}. \qquad \text{(A-4)}$$

For detection of a zero mean random signal S in additive independent zero mean noise N using a quadratic discriminant functional $X \mid C_1 = X_{SS} + X_{SN} + X_{NS} + X_{NN}$ and $X \mid C_2 = X_{NN}$, and ([2], [3], [29, Section 7.3], [35], [36], [38], [39], [47, Section 4.3], [49])

$$D = \frac{[m(X_{SS})]^2}{\sigma^2(X_{NN})}. \qquad \text{(A-4)}'$$

5) Complementary Deflection:

$$D^c = \frac{[m_1(X) - m_2(X)]^2}{\sigma_1^2(X)}. \qquad \text{(A-5)}$$

For the classification problem described in 4) [26], [45]

$$D^c = \frac{[m(X_{SS})]^2}{\sigma^2(X_{NN}) + \sigma^2(X_{SN} + X_{NS}) + \sigma^2(X_{SS})}. \qquad \text{(A-5)}'$$

6) Modified Deflection (for detection of a signal in additive noise N):

$$D^m \triangleq \frac{[m_1(X) - m_2(X)]^2}{\sigma_1^2(X - [X \mid N=0])}. \qquad \text{(A-6)}$$

For the classification problem described in 4) ([31])

$$D^m = \frac{[m(X_S)]^2}{\sigma^2(X_{NN}) + \sigma^2(X_{SN} + X_{NS})}. \qquad \text{(A-6)}'$$

7) Incremental SNR [32]:

$$\mathrm{ISNR} \triangleq \frac{[m_1(X)]^2 - [m_2(X)]^2}{\sigma_2^2(X)}. \qquad \text{(A-7)}$$

8) Differential SNR (for detection of a weak signal in additive noise $Y = \delta S + N$) [27]:

$$\mathrm{DSNR} \triangleq \frac{\left[\frac{d}{d\delta} m_1(X)\right]^2_{\delta=0}}{[\sigma_1^2(X)]_{\delta=0}} \tag{A-8}$$

where $\delta \mid C_1 > 0, \delta \mid C_2 = 0$.

9) Generalized SNR:

$$\mathrm{GSNR} \triangleq \frac{[m_1(X) - m_2(X)]^2}{w_1\sigma_1^2(X) + w_2\sigma_2^2(X)}, \tag{A-9}$$

where $w = \{w_1, w_2\}$ is a parameter vector of nonnegative weights ([16], [21], [24], [48]).

10) SNR [41]:

$$\mathrm{SNR} \triangleq \frac{[m_1(X) - m_2(X)]^2}{\sigma^2(X)}. \tag{A-10}$$

11) Generalized SNR′:

$$\mathrm{GSNR}' \triangleq \frac{[m_1(X) - m_2(X)]^2}{[\sigma_1^2(X)]^{w_1}[\sigma_2^2(X)]^{w_2}}. \tag{A-11}$$

Of these 11 second-order measures of quality, the majority are special cases of the remaining few, and/or are inappropriate in general, as briefly explained in this paragraph. Since the numerator of GSNR can be interpreted as a measure of the *separation* between the *centers* of the two class-conditional probability distributions of X, and the denominator can be interpreted as a measure of the *dispersion* of the two distributions away from their centers, then the ratio GSNR can be interpreted as a normalized measure of distance between the two distributions. The empirical version of GSNR with $w_i = p_i$ [see (15)-(17)] is Fisher's discrimination measure [15], [11, Section 4.10]. The two measures D and D^c are special cases of GSNR corresponding to the weight vectors $w = \{0, 1\}$, and $w = \{1, 0\}$, respectively. Neither of these is as appropriate as GSNR in general because each ignores one of the two dispersion terms. The measure D^m coincides with D^c when $\sigma^2(X_{SS}) = 0$ (e.g., S = phase randomized sinusoid, Φ = matched-filter-square-law-envelope-detector); in general, D^m ignores contributions to dispersion that can indeed affect discriminant performance. The measure ISNR coincides with D when $m_2(X) = 0$, and is not appropriate in general. The measure MSR does not distinguish between separation and dispersion. The measure SNR can be expressed in terms of GSNR (with weight vector $w = \{p_1, p_2\}$) by (19) and is, therefore, a monotonic strictly increasing function of GNSR. Thus, maximum GSNR and maximum SNR are equivalent design criteria. The measure ESNR coincides with SNR when $m_2(X) = 0$ [e.g., (A-1)′], but in general ESNR ignores the separation between centers. Although D, D^c, D^m, and DSNR are, in general, less appropriate than GSNR, all five of these measures of quality are approximately equal for the detection of a weak signal in additive noise (for which $\sigma_1 \cong \sigma_2$). The appropriateness of the measure MSE-SNR depends on the prescription of the signal parameter S. As shown in Section III, minimization of MSE-SNR for $S = \delta(C_1)$ (the class indicator) is equivalent to maximization of SNR. Other choices for S are discussed in a companion paper.[1] GSNR′ is an interesting alternative to GSNR that has not received attention in the literature. Another measure that is similar to GSNR, and is derived from Becker's measure of separability (cf. [6]) is

$$\mathrm{GSNR}'' \triangleq \frac{|m_1(X) - m_2(X)|}{\sigma_1(X) + \sigma_2(X)}. \tag{A-12}$$

B. M-ary Classification

As discussed in Section II-C, the most generally useful second order measures of quality for binary classification, MSE-SNR and SNR, can be applied to M-ary classification to obtain M-valued measures of quality: one value for each composite pair of classes, C_i and its complement "not C_i," and each component X_i of an M-tuple X of discriminants. As alternatives, various single-valued measures can be obtained by extension of the single-valued measures for binary classification. The second-order single-valued measures most commonly used are the scatter ratios which are defined in terms of the following scatter matrices.

1) *Between-Class Scatter:*

$$S_B \triangleq \sum_{i=1}^{M} p_i[m_i(X) - m(X)]\,[m_i(X) - m(X)]^t. \tag{A-13}$$

2) *Within-Class Scatter:*

$$S_W \triangleq \sum_{i=1}^{M} p_i k_i(X). \tag{A-14}$$

3) *Total Scatter:*

$$S_T \triangleq S_W + S_B \tag{A-15}$$

where $m_i(X)$ is the N-tuple and $k_i(X)$ the $N \times N$ matrix with elements defined by (7) and (8) respectively, and $[\cdot]^t$ denotes matrix transposition.

The most commonly used *scatter ratios* are defined as follows [17, Section 9.2]

$$J_1 \triangleq \mathrm{tr}\,\{S_W^{-1}S_B\}, \tag{A-16}$$

$$J_2 \triangleq \mathrm{tr}\,\{S_B\}/\mathrm{tr}\,\{S_W\}, \tag{A-17}$$

$$J_3 \triangleq \det\,\{S_B\}/\det\,\{S_W\} \tag{A-18}$$

where tr $\{\cdot\}$ and det $\{\cdot\}$ denote trace and determinant, respectively. In (A-18), it is assumed that $N = M$ so that det $\{S_B\} \neq 0$.) Alternative second-order measures of quality use S_T in place of S_W to obtain J_1', J_2', and J_3' from J_1, J_2, and J_3, respectively [17, Section 9.2]. Empirical measures are obtained by replacing p_i, m_i and k_i with $\tilde{p}_i$, $\tilde{m}_i$ and $\tilde{k}_i$ [(15)-(17)], respectively.

When $M = 2$ and $N = 1$, it follows from (A-13)-(A-18) that

$$J_1' = J_2' = J_3' = p_1 p_2 \text{ SNR} \tag{A-19}$$

and for $w = \{p_1, p_2\}$

$$J_1 = J_2 = J_3 = p_1 p_2 \text{ GSNR}. \tag{A-20}$$

Thus, each of the six scatter ratios is a formal extension of either GSNR or SNR from binary to M-ary classification. The relationships amongst J_1, J_2, and J_3 (and also, J_1', J_2' and J_3') for M-ary classification are discussed by Fukunaga [17, Section 9.2]. When $M > 2$ and $N > 2$, there is in general no relationship between the single-valued measures J_1, J_2, J_3, J_1', J_2' J_3' and the M-valued measures $\{\text{GSNR}_i(X_i)\}_1^M$, $\{\text{MSE}_i(X_i)\}_1^M$ and $\{\text{SNR}_i(X_i)\}_1^M$. However, when these measures are evaluated for a second-order-optimal generalized linear discriminant X_* (i.e., $X = X_*$ extremizes any, and, therefore, all of the three M-valued measures), then there is a simple relationship between one, and only one, of the single-valued measures (viz., J_1') and the three M-valued measures. This relationship is discussed in Sections II-C and III. Although these single-valued measures seem to be generally favored over the M-valued measures in the pattern recognition literature, the results in Section III indicate that the M-valued measures are more fundamental, and contain more information.

ACKNOWLEDGMENT

The author expresses his gratitude to Dr. R. S. Poulsen for his assistance with initial investigations of the relationship between the min-MSE and max-GNSR criteria [37].

REFERENCES

[1] V. R. Algazi, "Generalized SNR and performance index of filters for waveform estimation," *Proc. IEEE*, vol. 53, pp. 725–726, July 1965.

[2] C. R. Baker, "Optimum quadratic detection of a random vector in Gaussian noise," *IEEE Trans. Commun.*, vol. COM-14, pp. 802–805, Dec. 1966.

[3] ——, "On the deflection of a quadratic-linear test static," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 16–21, Jan. 1969.

[4] J. Capon, "Optimum weighting functions for the detection of sampled signals in noise," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 152–159, Apr. 1964.

[5] C. H. Chen, *Statistical Pattern Recognition*. Rochelle Park, NJ: Hayden, 1973.

[6] H. Chernoff, "The selection of effective attributes for deciding between hypotheses using linear discriminant functions," in *Proc. Int. Conf. Frontiers Pattern Recognition*, Honolulu, HI, Jan. 18–20, 1971, pp. 55–60.

[7] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21–27, Jan. 1967.

[8] P. A. Devijver, "Relationships between statistical risks and the least-mean-square-error design criterion in pattern recognition," in *Proc. 1st Int. Joint Conf. Pattern Recognition*, Washington, DC, Oct. 30–Nov. 1, 1973, pp. 139–148.

[9] ——, "On a new class of bounds on Bayes risk in multihypothesis pattern recognition," *IEEE Trans. Comput.*, vol. C-23, pp. 70–78, Jan. 1974.

[10] ——, "Decision-theoretic and related approaches to pattern classification," Section 8, ch. II, in *Pattern Recognition-Theory and Applications*, A. B. Whinston and K. S. Fu, Eds. Leydon: Sijthoff Int. Publ. Cy. (also, Rep. R308, MBLE Laboratoire de Recherches, Brussels, Belgium, Nov. 1975).

[11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[12] B. Epstein, *Linear Functional Analysis*. Philadelphia, PA: Saunders, 1970.

[13] P. L. Feintuch and N. J. Bershad, "Envelope detector performance for a partially coherent-fading sine wave in noise," *IEEE Trans. Commun.*, vol. COM-24, pp. 913–917, Aug. 1976.

[14] L. Fisher and J. W. Van Ness, "Admissable discriminant analysis," *J. Amer. Stat. Ass.*, vol. 68, pp. 603–607, Sept. 1973.

[15] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, Part II, pp. 179–188, 1936; also *Contributions to Mathematical Statistics*. New York: Wiley, 1950.

[16] K. Fukunaga and T. F. Krile, "A minimum distance feature effectiveness criterion," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 780–782, Sept. 1968.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

[18] K. Fukunaga and S. Ando, "The optimum nonlinear features for a scatter criterion in discriminant analysis," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 453–459, July 1977.

[19] W. A. Gardner, "The structure of least mean square linear estimators for synchronous M-ary signals," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 240–243, Mar. 1973.

[20] ——, "Structurally constrained receivers for signal detection and estimation," *IEEE Trans. Commun.*, vol. COM-24, pp. 578–592, June 1976. Erratta—the following errors in [20] should be noted: 1) interchange δ and $\{$ following (3), 2) change first + in (42) to −, 3) interchange "not H_i" and "not H_j" in (73), 4) replace m_x in (80) with closest X_i to m_x, 5) insert "finite" before "mean" preceding (A1), 6) replace $\delta(X/y)$ with $\delta(X)$ in (A2) and following (A4), and 7) change 535–544 to S34–S46 in [18].

[21] ——, "Noncoherent quadratic filtering for maximum SNR," in *Abstr. Papers Presented at IEEE Int. Symp. Inform. Theory*, Ronneby, Sweden, June 21–24, 1976, p. 144.

[22] ——, "Comparison of estimation-based classifiers," in *Abstr. Papers IEEE Int. Symp. Inform. Theory*, Ronneby, Sweden, June 21–24, 1976, p. 44.

[23] ——, "An equivalent linear model for marked and filtered doubly stochastic Poisson processes with application to MMSE linear estimation for synchronous M-ary optical data signals," *IEEE Trans. Commun.*, vol. COM-24, pp. 917–921, Aug. 1976.

[24] ——, "Anomalous behavior of receiver output SNR as a predictor of signal detection performance exemplified for quadratic receivers and incoherent fading Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 25, pp. 743–745, Nov. 1979.

[25] D. A. George, "Matched filters for interfering signals," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 153–154, Jan. 1965.

[26] M. Kanefsky, "Detection of weak signals with polarity coincidence arrays," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 260–268, Apr. 1966.

[27] S. A. Kassam and T. L. Lim, "Coefficient and data quantization in matched filters for detection," *IEEE Trans. Commun.*, vol. COM-26, pp. 124–127, Jan. 1978.

[28] J. S. Koford and G. F. Groner, "The use of an adaptive threshold element to design a linear optimal pattern classifier," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 42–50, Jan. 1966.

[29] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*. New York: McGraw-Hill, 1950.

[30] W. S. Meisel, *Computer-Oriented Approaches to Pattern Recognition*. New York: Academic, 1972.

[31] D. Middleton, "Acoustical signal detection by simple correlators in the presence of non-Gaussian noise. I. Signal-to-noise ratios and canonical forms," *J. Acoust. Soc. Amer.*, vol. 34, pp. 1598–1609, Oct. 1962.

[32] J. W. Modestino, "Adaptive detection of signals in impulsive noise environments," *IEEE Trans. Commun.*, vol. COM-25, pp. 1022–1027, Sept. 1977.

[33] N. J. Nilsson, *Learning Machines*. New York: McGraw-Hill, 1965.

[34] J. D. Patterson and B. F. Womack, "An adaptive pattern classification system," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-2, pp. 62–67, Aug. 1966.

[35] B. Picinbono, "Optimum filtering and multichannel receivers," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 256–260, Apr. 1966.

[36] H. V. Poor, "A robustness property of signal-to-noise ratios," in *Proc. 12th Ann. Conf. Inform. Sci. Syst.*, Mar. 1978.

[37] R. S. Poulsen, "Evaluation of the constrained Bayesian methodology for signal detection," Ph.D. dissertation, Dep. Elec. Eng., Univ. California, Davis, Dec. 1977.

[38] R. Price and P. E. Green, Jr., "Signal processing in radar astronomy—Communication via fluctuating multipath media," Lincoln Lab., M.I.T., Tech. Rep. 234, Oct. 1960.

[39] R. Price, "Output signal-to-noise ratio as a criterion in spread-channel signaling," Lincoln Lab., M.I.T. Tech. Rep. 388, May 1965.

[40] J. Rubin, "Optimal classification into groups: An approach for solving the taxonomy problem," *J. Theoretical Biol.*, vol. 15, pp. 103–44, 1967.

[41] P. Rudnick, "A signal-to-noise property of binary decisions," *Nature*, vol. 193, pp. 604–605, Feb. 1962.

[42] G. S. Sebestyen, *Decision Making Processes in Pattern Recognition.* New York: Macmillan, 1962.

[43] S. E. Smith and S. S. Yau, "Linear sequential pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 673–678, Sept. 1972.

[44] A. Som and A. K. Nath, "Sequential pattern classifier using least-mean-square-error criterion," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, pp. 439-443, July 1972.

[45] G. L. Turin, "An introduction to digitial matched filters," *Proc. IEEE*, vol. 64, pp. 1092–1112, July 1976.

[46] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

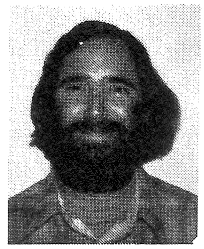
[47] ——, *Detection, Estimation, and Modulation Theory, Part III.* New York: Wiley, 1971.

[48] A. N. Venetsanopoulos and F. B. Tuteur, "Transmitter-receiver optimization for active signaling over undersea acoustic channels," *IEEE Trans. Commun.*, vol. COM-19, pp. 649–659, Oct. 1971.

[49] ——, "Upper bound on the signal-to-noise ratio at the output of a threshold receiver," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 753–755, Nov. 1971.

[50] W. G. Wee, "Generalized inverse approach to adaptive multiclass pattern classification," *IEEE Trans. Comput.*, vol. C-17, pp. 1157–1164, Dec. 1968.

★



William A. Gardner (S'64–M'67) was born in Palo Alto, CA, on November 4, 1942. He received the M.S. degree from Stanford University, Stanford, CA, in 1967, and the Ph.D. degree from the University of Massachusetts, Amherst, in 1972, both in electrical engineering.

He was a Member of the Technical Staff at Bell Laboratories from 1967 to 1969. He has been a Faculty Member in the Department of Electrical and Computer Engineering, University of California, Davis, since 1972, where he is currently an Associate Professor. His research interests are generally in statistical communication theory and random processes, and primarily in detection and estimation theory.

Dr. Gardner is a member of Alpha Gamma Sigma, Eta Kappa Nu, Sigma Xi, and Tau Beta Pi.