Structural Characterization of Locally Optimum Detectors in Terms of Locally Optimum Estimators and Correlators

WILLIAM A. GARDNER, MEMBER, IEEE

Abstract—Explicit formulas for locally (SNR $\to$ 0) optimum (MMSE) signal estimators (smoother, filter, and predictor) for discrete-time observations of a random signal in additive random noise are derived and used to characterize the locally optimum (likelihood ratio) signal detector for on–off signaling. The characterizations are canonical (distribution-free) detector structures involving estimator–correlators. These structural characterizations provide new interpretations of known detectors for various special cases. If the one-step signal predictor is recursive and the noise is white (possibly non-Gaussian or nonstationary), there is a canonical structure that admits recursive computation. The primary motivation for these structural characterizations is to render the estimator–correlator design philosophy applicable for the purpose of simplifying implementations and enhancing adaptability. Unlike the known estimator–correlator structural characterizations for continuous-time globally optimum detectors, the new characterizations apply for non-Gaussian as well as Gaussian noise, and the estimators are explicit rather than implicit.

I. INTRODUCTION

A. Motivation

THE practical value of the signal-detector design philosophy based on an estimator–correlator structural characterization of the optimum detector, for a Gaussian signal in additive Gaussian noise, was demonstrated [1] shortly after it was proposed [2], 25 years ago. It is shown in [1] that the filter in the optimum detector can be significantly simplified (for the purpose of implementation) with only negligible degradation in detection performance. As demonstrated more recently [3], the practical value of simplified implementation afforded by this characterization can be exploited in a variety of ways while maintaining near optimal performance. The practical value of an analogous design philosophy for point and jump process signal detection [4] also has been demonstrated by elimination of nonlinearities with negligible performance degradation [5]. Based on this practical motivation of simplified implementation, a number of investigators have contributed to the derivation of estimator–correlator structural characterizations, and, more generally, the elucidation of the role of signal estimation in optimum (likelihood ratio) detection for a wide variety of probabilistic models for continuous-time observations [6, and references therein], [7]–[11]. However, efforts to develop analogous characterizations for probabilistic models for discrete-time observations have met with more limited success for all but the most elementary model of a Gaussian signal in additive Gaussian noise [12, and references therein]. This is illustrated by the fact that the estimator–correlator characterization for continuous-time observation of a non-Gaussian signal in additive Gaussian noise has no counterpart for discrete-time observation [12, and references therein], except in a generalized sense [13] that has not been shown to be of practical value as a basis for a design philosophy (but does yield a recursive formula, which might suggest an efficient implementation).

Motivated by the increasing practical importance of discrete-time implementation and of low-SNR detection [29], the purpose of this paper is to show that, in contrast to the limited applicability of the estimator–correlator design philosophy to globally (arbitrary SNR) optimum detection with discrete-time observations, this philosophy is widely applicable to locally (SNR $\to$ 0) optimum detection with discrete-time observations.[1]

It is well known that the structure of the locally optimum detector for an arbitrary on–off random signal in additive arbitrary noise is mathematically explicit [14], and that for several special cases (such as coherent and noncoherent detection of a sine wave in white noise [15], [16], [29]) the structure is functionally explicit, consisting of functional elements such as zero-memory nonlinearities, correlators, squarers, and summers. For these special cases, the applicability of the estimator–correlator design philosophy is based on alternative structural characterizations of known structures. For some of these special cases, the known structure is sufficiently amenable to implementation that the design philosophy based on a structural reinterpretation in terms of an estimator–correlator is unlikely to lead to a simplified implementation (although it can enhance adaptability). However, the structural char-

[1]At the time of preparation of this manuscript, the author was not aware of other work on this topic. However, a literature survey conducted before submission of this manuscript for publication revealed the work of Sosulin [27] (and references therein) on locally optimum recursive estimator–correlator structures for on–off Markov signals in additive Markov noise and some generalizations thereof. The results in [27] are, in essence, embellishments of the structural characterization (33) derived in Section V. This derivation, however, applies to a general signal model, whereas Sosulin's applies to a Markov signal model.

Manuscript received February 4, 1981; revised April 30, 1981.
The author is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616.

0018-9448/82/1100-0924$00.75 ©1982 IEEE

acterization can still be of conceptual value in such cases. For example, for detection of a sine wave transmitted over a Rician channel [17, p. 360], the characterization reveals that it is the sum of the specular signal component and the MMSE (linear) estimate of the Rayleigh fading component that is correlated with the observations; and, regardless of the fading distribution and the noise distribution, the characterization (Section IV) reveals that the signal to be correlated with is still the sum of the specular component and the local-MMSE (nonlinear) estimate of the fading component.

As another example of practical value, the characterization reveals that the only difference between coherent and noncoherent detectors for an arbitrary cyclostationary signal (AM, PM, PAM, PCM, etc.) in additive arbitrary noise is that a synchronized periodically time-varying linear estimation filter in the coherent detector is replaced with its time-averaged counterpart in the noncoherent detector. The practical significance of this is discussed in Section VI, and in more detail in [3]. This simple relationship between coherent and noncoherent detectors applies to *all* weak cyclostationary signal in additive noise detection problems, whereas, previously known explicit relationships between coherent and noncoherent detectors require that the signal be narrow-band, the noise be broad-band, the signal phase and amplitude be independent, the noise distribution be circularly symmetric, etc. [29, and references therein].

For some signal detection problems (such as detection of a random signal in non-Gaussian narrow-band noise [16]), the locally optimum detector can be sufficiently complex that the design philosophy based on the estimator–correlator structural characterization can be of significant practical value in obtaining a simplified implementation, especially if adaptation is required. Since the particular noise (or signal) distribution governing observations in a given application is likely to be unknown and possibly time-varying, an adaptive implementation of a locally optimum detector is often indicated [15], [16], [18]. In such a signal/noise environment, the estimator–correlator structural characterization can be used to advantage. For example, for serial signal detection (e.g., digital data transmission), a decision-directed adaptive estimator can be used (possibly following an initial supervised training mode). One advantage offered by the estimator–correlator structure for adaptive implementation is that a performance indicator for control of the adaptation is available during operation; viz., the error between the decided (detected) signal and the estimated signal. (This is an alternative to the SNR performance measure employed in [18].)

As additional motivation, it is mentioned that in addition to being near optimal for low SNR, the locally optimum detector can yield adequate performance for larger SNR [15], [16]. Also, the locally optimum detector can be near optimal under conditions of low-energy coherence in the signal, even when the detector output SNR is large (yielding high-performance detection) [3]. Therefore, estimator–correlator structural characterizations of locally optimum detectors can be of practical value for moderate-to-high performance detection and for moderate-to-low noise, as well as for low-performance detection and high noise. The performance of locally optimum detectors, according to various criteria, has received considerable attention [14]–[16], [18]–[22], [29]. For example, certain locally optimum detectors have been shown to be asymptotically (observation time $\to \infty$) optimum.

B. Problem Statement

The class of detection problems to be considered is modeled by the hypothesis testing problem

$$H_0\colon\ y_i = n_i,$$
$$H_1\colon\ y_i = s_i + n_i \qquad i = 1, 2, \cdots, m,$$

for which the m observations $y = \{y_i\}_1^m$ consist of an m-vector of random noise variables (with joint probability density (pdf) denoted by $f_N(\boldsymbol{n})$) under the null hypothesis H_0, and consist of the sum of the noise m-vector and an m-vector of random signal variables (with joint pdf denoted by $f_S(s)$) under the alternative hypothesis H_1. It is assumed that the signal vector $\boldsymbol{s}$ is statistically independent of the noise vector $\boldsymbol{n}$. Multiplicative noise can be included in the model for $\boldsymbol{s}$, as described in Section VII.

As is well known, the optimum (Bayes, Neyman–Pearson, or minimax) signal detector is (equivalent to) the log likelihood ratio test

$$\tau(y) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

for some threshold value γ, where

$$\tau(y) \triangleq \log\left[f_{Y|H_1}(y)\right] - \log\left[f_{Y|H_0}(y)\right]. \tag{1}$$

The conditional densities in (1) are given by

$$f_{Y|H_0}(y) = f_N(y),$$
$$f_{Y|H_1}(y) = \int f_N(y - s) f_S(s)\, ds, \tag{2}$$

where the integral is m-fold. When the signal is known, f_S is an m-dimensional Dirac delta. For later reference, we introduce the notation

$$g(y) \triangleq \log\left[f_N(y)\right],$$
$$g_i(y_i) \triangleq \log\left[f_{N_i}(y_i)\right]. \tag{3}$$

The class of estimation problems to be considered is described by H_1. As is well known, the optimum (MMSE) signal estimator is the conditional mean; e.g., for the smoothing estimator,[2]

$$\hat{s}(y) = E\{S\,|\,y\} = \int s f_{S|Y}(s\,|\,y)\, ds. \tag{4}$$

The objective in Sections II and III is to obtain approximations to the estimation function, $\hat{s}(\cdot)$ (for smoothing, filtering, and prediction) and the detection function, $\tau(\cdot)$, that are asymptotically (SNR $\to 0$) exact. These are

[2]We use capital letters for random m-vectors, and corresponding lowercase letters for realizations (statistical samples) of the random m-vectors.

the *locally optimum estimator and detector*. The objective in Sections IV and V is to characterize $\tau(\cdot)$ in terms of $\hat{s}(\cdot)$.

To define the statement SNR $\to 0$, we denote the random signal vector S by

$$S \triangleq \delta S_0,$$

where δ is a nonrandom scalar, and S_0 is any "normalized" version (e.g., trace of correlation matrix for S_0 is unity) of the m-vector S. Then

$$\mathrm{SNR} \to 0 \Leftrightarrow \delta \to 0. \tag{5}$$

C. Overview

In Section II, explicit formulas for the locally optimum (MMSE) noncausal fixed-interval smoother and the causal filter and one-step predictor are derived. It is shown that these causal estimators are recursive if the signal is wide-sense Markov and the noise is white, but not necessarily Gaussian or stationary. In addition, it is shown that the estimator is linear if the noise is Gaussian.

In Section III, a derivation of the well-known mathematically explicit formula for the local log likelihood ratio is presented for completeness.

In Section IV, the formulas from Sections II and III are employed to derive a canonical (distribution-free) estimator–correlator structural characterization of the local log likelihood ratio. It is shown that this structure admits a recursive implementation if the signal is wide-sense Markov and the noise is white, but not necessarily Gaussian or stationary. In addition, it is shown that this structure decomposes into a linear estimator–correlator preceded by a noise whitener if the noise is Gaussian.

In Section V, an alternative derivation of the locally optimum detector that is tailored to the objective of recursive computation is presented. It is shown that the local log likelihood ratio can be computed recursively if the noise is white and the one-step signal predictor is recursive.

In Section VI, the estimator–correlator structural characterization is employed to compare coherent and noncoherent detector structures for arbitrary cyclostationary signals in additive arbitrary noise.

Finally, in Section VII, applicability to multiplicative noise is briefly discussed, and potential extension to jump and point process observations is suggested.

II. Locally Optimum Estimators

By using Bayes law, (4) can be expressed as

$$\hat{s}(y) = \frac{\int s f_{Y|S}(y \mid s) f_S(s)\, ds}{\int f_{Y|S}(y \mid s) f_S(s)\, ds}, \tag{6}$$

where

$$f_{Y|S}(y \mid s) = f_N(y - s). \tag{7}$$

The first-order Taylor series expansion of the function f_N about the point y, evaluated at $y - s$ is[3]

$$f_N(y - s) \simeq f_N(y) - s^T \nabla f_N(y), \tag{8}$$

where $\nabla f_N(y)$ is the gradient vector of f_N evaluated at y and has ith element $[\nabla f_N(y)]_i = \partial f_N(y)/\partial y_i$. It follows from (3) that

$$\nabla f_N(y)/f_N(y) = \nabla g(y). \tag{9}$$

Substitution of (8) into (7) into (6), and use of (9) yields the approximation

$$\hat{s}(y) \simeq \frac{\bar{s}[1 - \bar{s}^T \nabla g(y)] - K_S \nabla g(y)}{1 - \bar{s}^T \nabla g(y)}, \tag{10}$$

which is further approximated (by ignoring the vanishingly small term $\bar{s}^T \nabla g(y)$ for SNR $\to 0$) by

$$\hat{s}(y) \simeq \bar{s} + K_S[-\nabla g(y)], \tag{11}$$

where K_S is the covariance matrix for S,

$$K_S = E\{SS^T\} - \bar{s}\bar{s}^T,$$

and $\bar{s}$ is the mean vector for S,

$$\bar{s} = E\{S\}.$$

For example, $\bar{s}$ can be interpreted as the known signal component, so that K_S is the autocorrelation matrix for the random signal component, $S - \bar{s}$; $K_S = E\{(S - \bar{s})(S - \bar{s})^T\}$. Eq. (11) can be written out as

$$\hat{s}_i(y) \simeq \bar{s}_i + \sum_{j=1}^{m} K_S(i, j)[-\partial g(y)/\partial y_j], \tag{11$'$}$$

where $K_S(i, j) = E\{S_i S_j\} - \bar{s}_i \bar{s}_j$.

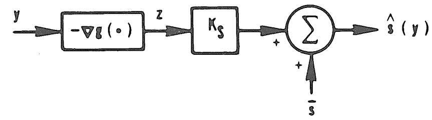


Fig. 1. Local-MMSE estimator.

A signal flow block diagram of this local-MMSE estimator is shown in Fig. 1. The observations y are transformed by the nonlinear (in general) transformation $-\nabla g(\cdot)$, which is specified by the noise pdf. Then the transformed observations z are filtered (matrix product) by the linear transformation K_S, which is specified by the signal pdf. Finally, the known signal component $\bar{s}$ is added. The nonlinearity, $-\nabla g(\cdot)$, has been studied extensively (in connection with signal detection) for various non-Gaussian noise distributions that arise in practice [15], [16], [29].

The estimator (6) and (11) is the *noncausal* (*smoothing*) *estimator* for s_i in terms of the observations $y = \{y_j: 1 \le j \le m\} \triangleq y^m$. By letting $m = i$ in (11$'$) with y denoted by y^m, (11$'$) yields the *causal* (*filtering*) *estimator*

$$\hat{s}_i(y^i) \simeq \bar{s}_i + \sum_{j=1}^{i} K_S(i, j)[-\partial g(y^i)/\partial y_j]. \tag{12}$$

Similarly, the *one-step predicting estimator* can be shown to

[3] The superscript T denotes matrix transposition; thus, since s is taken to be a column vector s^T is a row vector.

be

$$\hat{s}_{i+1}(y^i) \simeq \bar{s}_{i+1} + \sum_{j=1}^{i} K_S(i+1, j)\left[-\partial g(y^i)/\partial y_j\right]. \tag{13}$$

A. Independent Noise and Recursive Estimation

If the noise variables $\{N_i\}$ are independent, then

$$-\partial g(y^i)/\partial y_j = -dg_j(y_j)/dy_j \triangleq z_j(y_j).$$

Therefore, both the filtering estimator (12) and the predicting estimator (13) can be computed recursively provided that the covariance K_S exhibits Markov structure. For example, the first-order Markov structure $K_S(i, j) = \sigma_S^2 r^{|i-j|}$ makes (12) reduce (with $\bar{s} = \mathbf{0}$ for simplicity) to the first-order recursion

$$\hat{s}_i(y^i) = r\hat{s}_{i-1}(y^{i-1}) + \sigma_S^2 z_i(y_i).$$

Furthermore, in this case, the one-step predictor and the filtered estimator are related by (letting $\bar{s} = \mathbf{0}$ for simplicity)

$$\hat{s}_{i+1}(y^i) = r\hat{s}_i(y^i),$$
$$\hat{s}_i(y^i) = \hat{s}_i(y^{i-1}) + \sigma_S^2 z_i(y_i).$$

B. Gaussian Noise

In general, these local-MMSE estimators (11)–(13) are nonlinear. However, if the noise is Gaussian, then

$$g(y) = -\frac{1}{2}(y-\bar{n})^T K_N^{-1}(y-\bar{n}) + \text{constant}, \tag{14}$$

and therefore $\nabla g(y) = -K_N^{-1}(y-\bar{n})$, from which (11) (for example) reduces to

$$\hat{s}(y) \simeq \bar{s} + K_S K_N^{-1}(y-\bar{n}).$$

This is clearly an approximation (SNR $\to 0$) to the *linear* MMSE estimator

$$\hat{s}_L(y) = \bar{s} + K_S[K_S + K_N]^{-1}(y - \bar{s} - \bar{n}).$$

III. Locally Optimum Detector

The log likelihood ratio is, from (1) and (2),

$$\tau(y) = \log \int f_N(y-s) f_S(s)\, ds - \log f_N(y). \tag{1'}$$

The second-order Taylor series expansion[4] of the function f_N about the point y, evaluated at $y - s$ is

$$f_N(y-s) \simeq f_N(y) - s^T \nabla f_N(y) + \frac{1}{2} s^T H_{f_N}(y) s, \tag{15}$$

where $H_{f_N}(y)$ is the Hessian matrix of f_N evaluated at y, and has (ij)th element

$$\left[H_{f_N}(y)\right]_{ij} = \partial^2 f_N(y)/\partial y_i \partial y_j.$$

[4]Although only two terms were retained in Section II (8), we must retain three terms here (15) because the first two terms, when substituted in (1′), represent only the signal mean, and not the random component of the signal.

It follows from (3) that

$$H_{f_N}(y)/f_N(y) = H_g(y) + \nabla g(y)[\nabla g(y)]^T. \tag{16}$$

Substitution of (15) into (1′), and use of (9) and (16) yields[5]

$$\tau(y) \simeq \log\left[1 - \bar{s}^T \nabla g(y)\left[1 - \frac{1}{2}\bar{s}^T \nabla g(y)\right] + \frac{1}{2}[\nabla g(y)]^T K_S \nabla g(y) + \frac{1}{2}\operatorname{tr}\left\{\left[K_S + \bar{s}\bar{s}^T\right] H_g(y)\right\}\right], \tag{17}$$

which is further approximated (by use of $1 - \frac{1}{2}\bar{s}^T \nabla g(y) \simeq 1$ for SNR $\to 0$, and by use of $\log(1-x) \simeq -x$, since $x \to 0$ as SNR $\to 0$) by

$$\tau(y) \simeq -\bar{s}^T \nabla g(y) + \frac{1}{2}[\nabla g(y)]^T K_S \nabla g(y) + \frac{1}{2}\operatorname{tr}\left\{\left[K_S + \bar{s}\bar{s}^T\right] H_g(y)\right\}. \tag{18}$$

This formula is identical to Middleton's formula [14, eq. (19)].

IV. Estimator–Correlator Structures

In order to characterize the local log likelihood ratio in terms of the local-MMSE estimator, we first manipulate (18) into the appropriate form. This requires the following characterization of the Hessian operator (on $g(y)$) as the outer product of the gradient operator (on $g(y)$) with itself,

$$H_g(y) = \nabla \nabla^T g(y). \tag{19}$$

In the remainder of this section, approximations (11)–(13) and (18) will be written as equalities, which simply means that $\tau(\cdot)$ and $\hat{s}(\cdot)$ are to be interpreted as locally optimum, rather than globally optimum. Substitution of (19) into (18), and regrouping of terms involving $\bar{s}$ yields

$$\tau(y) = \frac{1}{2}\bar{s}^T[-\nabla g(y)] + \frac{1}{2}(\bar{s} + K_S[-\nabla g(y)])^T[-\nabla g(y)] - \frac{1}{2}\operatorname{tr}\left\{\nabla(\bar{s} + K_S[-\nabla g(y)])^T + \nabla(\bar{s}^T[-\nabla g(y)]\bar{s}^T)\right\}. \tag{18'}$$

Substitution of (11) into (18′) yields the desired characterization:

$$\tau(y) = \frac{1}{2}[\bar{s} + \hat{s}(y)]^T[-\nabla g(y)] - \frac{1}{2}\operatorname{tr}\left\{\nabla[\hat{s}(y)]^T + \nabla\left[\bar{s}^T[-\nabla g(y)]\bar{s}^T\right]\right\} \tag{20}$$

$$\tau(y) = \sum_i \frac{1}{2}(\bar{s}_i + \hat{s}_i(y))z_i - \sum_i \frac{\partial}{\partial y_i}\left[\hat{s}_i(y) + \left(\sum_j \bar{s}_j z_j\right)\bar{s}_i\right]\frac{1}{2}, \tag{20'}$$

[5]The quantity $\operatorname{tr}\{M\}$ denotes the trace of the square matrix M.

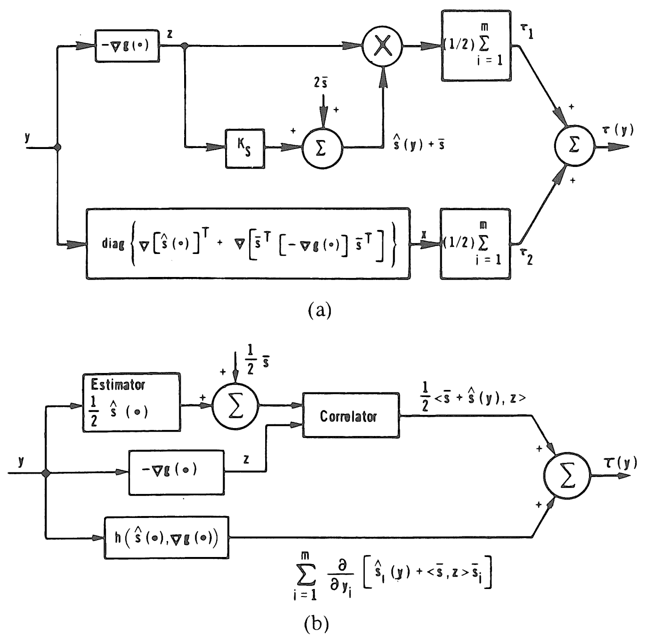


Fig. 2. (a) Locally optimum canonical detector structure, characterized in terms of the local-MMSE estimator (smoother). (b) Alternative schematic for the detector structure in Fig. 2(a)

where

$$z \triangleq -\nabla g(y).$$

This local log likelihood ratio formula can be represented by the signal-flow block diagram shown in Fig. 2(a), where the alternative notation[6]

$$\operatorname{tr}\left\{\nabla[\hat{s}(y)]^T\right\} = \sum_{i=1}^{m}\left[\operatorname{diag}\left\{\nabla[\hat{s}(y)]^T\right\}\right]_i$$

is used. The upper signal-flow path in this canonical structure, which corresponds to the first sum in (20′), is an estimator–correlator. That is, the observations y are first transformed by $-\nabla g(\cdot)$ into z. Then the local-MMSE estimate $\hat{s}(y)$ is obtained from z by the filter (matrix) K_S and addition of the known component $\bar{s}$. Finally, the arithmetic mean of the known component and the estimated signal is correlated with the transformed observations; i.e., the discrete-time signals $[\hat{s}(y) + \bar{s}]/2$ and z are multiplied together and summed (over time). In the special case for which the signal is completely known ($K_S \equiv 0$), the upper signal-flow path degenerates into the conventional locally optimum correlator detector for a known signal [15, and refs. therein]. In the special case for which the signal is completely random ($\bar{s} \equiv 0$), the upper signal-flow path remains an estimator–correlator and is a new interpretation of the conventional locally optimum quadrature correlator detector for various applications involving bandpass signals with random amplitude and phase [16, and refs. therein].

The lower signal-flow path in the canonical structure, which corresponds to the second sum in (20′), involves the derivatives, with respect to the observations, of the signal estimator and the known signal-component part of the correlation performed in the upper path. In certain special cases, this signal-flow path degenerates to yield a term that is independent of the observations, but that contributes an important bias to the threshold—an SNR-dependent bias. These and other special cases are discussed in the following subsections.

For application of the estimator–correlator design philosophy, the alternative signal-flow block diagram shown in Fig. 2(b) should be used. The only functional elements contained in the upper signal-flow path of this diagram are the signal estimator $\hat{s}(\cdot)$, a nonlinearity, $\nabla g(\cdot)$, which is dependent on only the noise distribution, and a correlator, $\langle \cdot , \cdot \rangle$. The lower signal-flow path can be implemented using only the "hardware" in the upper signal-flow path, but in a time-shared mode, provided that the derivatives $\partial/\partial y_i$ are approximated by finite differences.

A. Gaussian Noise

If the noise is Gaussian, then it follows from (14) that $H_g(y) = -K_N^{-1}$, in which case the output of the lower signal-flow path in Fig. 2(a) is simply the signal-independent bias term

$$\tau_2 = -\operatorname{tr}\left\{K_S K_N^{-1}\right\} - \bar{s}^T K_N^{-1}\bar{s}. \tag{21}$$

This term is closely related (in fact, equal in some special cases) to the SNR-dependent bias term in the optimum detectors for known and Gaussian signals in Gaussian noise [23, ch. 2]. Furthermore, for this special case of Gaussian noise, the nonlinear transformation in the upper signal-flow path degenerates into the linear transformation $-\nabla g(y) = K_N^{-1}(y - \bar{n})$, which removes the mean noise, and then whitens the remaining zero-mean noise. Specifically, (20′) reduces to (letting $\bar{s} = \bar{n} = 0$ for simplicity)

$$\tau(y) = \frac{1}{2}[\hat{s}(y)]^T K_N^{-1} y - \frac{1}{2}\operatorname{tr}\left\{K_S K_N^{-1}\right\}, \tag{22}$$

$$\tau(y) = \frac{1}{2}[\hat{\tilde{s}}(\tilde{y})]^T \tilde{y} - \frac{1}{2}\operatorname{tr}\left\{K_{\tilde{S}}\right\} \tag{22'}$$

where

$$\tilde{y} = K_N^{-1/2} y, \quad \tilde{y} \mid H_1 = \tilde{s} + \tilde{n}, \tag{23}$$

and

$$K_{\tilde{S}} = K_N^{-1/2} K_S K_N^{-1/2}, \quad K_{\tilde{N}} = I. \tag{24}$$

In (23), $K_N^{-1/2}$ is the unique, positive definite, symmetric square root of the matrix K_N (which is assumed to be positive definite). The signal-flow block diagram for this locally optimum detector for zero-mean random signals in additive zero-mean Gaussian noise is shown in Fig. 3. This noise whitener estimator–correlator structure is functionally similar to the optimum detector structure for continuous-time observations [6], however, the local-MMSE estimator in Fig. 3 is noncausal, whereas that in the optimum detector (for non-Gaussian signals) is causal, and similarly for the noise whitener. Nevertheless, both noncausal filters $K_{\tilde{S}}$ and $K_N^{-1/2}$, can be replaced with equivalent causal filters. Specifically, by defining $K_N^{-1/2}$ to be the causal

[6]The quantity $\operatorname{diag}\{M\}$ is the m-vector on the diagonal of the $m \times m$ matrix M.

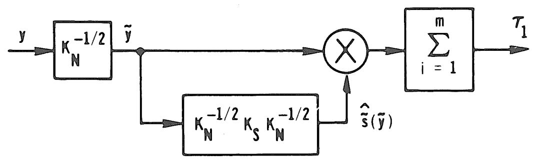


Fig. 3. Locally optimum detector for zero-mean signal in additive Gaussian noise: Local-MMSE estimator–correlator preceded by a noise whitener.

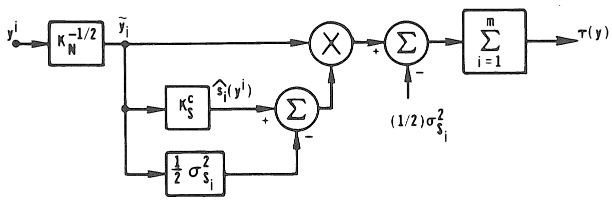


Fig. 4. Optimum causal noise-whitener causal estimator–correlator detector structure for a zero-mean signal in additive Gaussian noise.

square root of K_N^{-1}, and by using the identity

$$\tilde{y}^T K_{\tilde{S}}^c \tilde{y} = \frac{1}{2}\tilde{y}^T K_{\tilde{S}} \tilde{y} + \frac{1}{2}\sum_{i=1}^{m}\left(\sigma_{\tilde{S}_i}\tilde{y}_i\right)^2,$$

where $K_{\tilde{S}}^c$ is the causal part of $K_{\tilde{S}}$, i.e.,

$$K_{\tilde{S}}^c(i,j) = \begin{cases} K_{\tilde{S}}(i,j), & j \le i, \\ 0, & j > i, \end{cases}$$

(22′) can be reexpressed as

$$\tau(y) = \sum_{i=1}^{m} \hat{\tilde{s}}_i(\tilde{y}^i)\tilde{y}_i - \frac{1}{2}\sum_{i=1}^{m}\sigma_{\tilde{S}_i}^2\left(1+\tilde{y}_i^2\right), \qquad (22'')$$

where $\hat{\tilde{s}}_i(\tilde{y}^i)$ is the causal estimator for $\tilde{s}_i$ (12), and $K_N^{-1/2}$ in (23) is the causal noise whitener. The signal-flow block diagram for this locally optimum causal detector is shown in Fig. 4.

B. Independent Noise

Another special case of interest is that for which the noise variables are independent. In this case

$$[\nabla g(y)]_i = dg_i(y_i)/dy_i,$$
$$[H_g(y)]_{ij} = \left(d^2 g_i(y_i)/dy_i^2\right)\delta_{ij}, \qquad (25)$$

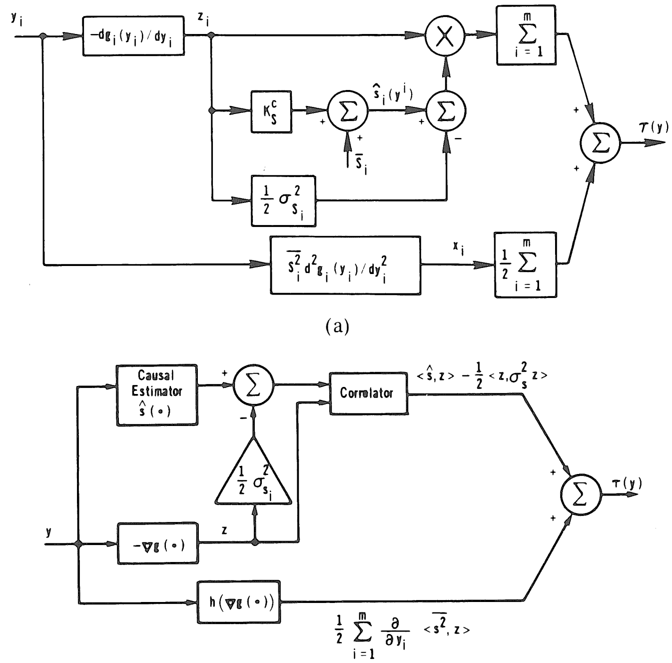
where δ_{ij} is the Kronecker delta. As a result of (25), the nonlinear transformation $-\nabla g(\cdot)$ in the upper signal-flow path in Fig. 2(a) is memoryless (and therefore causal), and the nonlinear transformation in the lower signal-flow path is also memoryless, i.e., the ith terms at the outputs of these transformations are

$$z_i(y_i) = -dg_i(y_i)/dy_i,$$
$$x_i(y_i) = \left(\sigma_{S_i}^2 + \bar{s}_i^2\right)d^2 g_i(y_i)/dy_i^2. \qquad (26)$$

Therefore, by using the identity

$$z^T K_S^c z = \frac{1}{2} z^T K_S z + \frac{1}{2}\sum_{i=1}^{m}\left(\sigma_{S_i} z_i\right)^2$$

together with (26), (20′) admits a characterization in terms



(a)

(b)

Fig. 5. (a) Locally optimum causal estimator–correlator detector structure for independent noise variables. (b) Alternative schematic for the detector structure in Fig. 5(a).

of the causal estimator $\hat{s}_i(y^i)$ (12),

$$\hat{s}_i(y^i) = \bar{s}_i + \sum_{j=1}^{i} K_S(i,j) z_j(y_j),$$

viz.,

$$\tau(y) = \sum_{i=1}^{m}\hat{s}_i(y^i) z_i + \frac{1}{2}\sum_{i=1}^{m}\left[x_i(y_i) - \left(\sigma_{S_i} z_i(y_i)\right)^2\right]. \qquad (27)$$

A signal-flow block diagram for this locally optimum causal detector is shown in Fig. 5(a). We note that if K_S exhibits Markov structure then $\hat{s}_i(y^i)$, and therefore $\tau(y)$, can be computed recursively as discussed in Section II.

For application of the estimator–correlator design philosophy, the alternative signal-flow block diagram shown in Fig. 5(b) should be used. The only functional elements contained in the upper signal-flow path of this diagram are the causal signal estimator $\hat{s}(\cdot)$, a scaling amplifier with gain $\sigma_{S_i}^2$, a zero-memory nonlinearity, $\nabla g(\cdot)$, which is dependent on only the noise distribution, and a correlator, $\langle\cdot,\cdot\rangle$. The lower signal-flow path can be implemented using only the nonlinearity $\nabla g(\cdot)$, provided that the derivatives, $\partial/\partial y_i$, are approximated by finite differences, and $\nabla g(\cdot)$ is implemented in a time-shared mode.

If the independent noise variables are Laplacian,

$$f_{N_i}(n_i) = \left(\sqrt{2}\,\sigma_{N_i}\right)^{-1}\exp\left(-\sqrt{2}\,|n_i|/\sigma_{N_i}\right),$$

then (26) reduces to

$$z_i(y_i) = \frac{2}{N_i}\,\mathrm{sgn}(y_i),$$
$$x_i(y_i) = 0,$$

and therefore, the lower signal-flow path in Figs. 2 and 5 vanishes, and the nonlinearity in the upper path is simply a time-varying clipper. This, of course, includes the well-known locally optimum clipper-correlator structure for known signals, and the locally optimum clipper quadrature correlator structure for incoherent bandpass signals as special cases of these locally optimum clipper-estimator-correlator structures.

In general, if the function g is not nearly quadratic or lower order (piecewise linear)—i.e., if the noise is not nearly Gaussian or Laplacian—then the lower signal-flow path in the canonical structures cannot be deleted without incurring suboptimality.

V. Recursive Estimator–Correlator Structure

It follows from (27) that if the noise is independent, then the detection statistic can be computed recursively provided that the causal estimator $\hat{s}_i(y^i)$ can be computed recursively. An alternative approach to the derivation of locally optimum detectors that is tailored to the objective of recursive computation is presented in this section.

By using Bayes law, the likelihood ratio $L(y^i)$, which is based on observations up to time i, can be expressed in terms of $L(y^{i-1})$ as follows

$$L(y^i) = L(y^{i-1})\left[\frac{f_{Y_i|Y^{i-1},H_1}(y_i|y^{i-1})}{f_{Y_i|Y^{i-1},H_0}(y_i|y^{i-1})}\right].$$

We shall assume that the noise variables are independent so that

$$f_{Y_i|Y^{i-1},H_0}(y_i|y^{i-1}) = f_{Y_i|H_0}(y_i) = f_{N_i}(y_i). \tag{28}$$

By again using Bayes's law, we obtain the expression

$$f_{Y_i|Y^{i-1},H_1}(y_i|y^{i-1}) = \int f_{Y_i|Y^{i-1},S_i}(y_i|y^{i-1},s_i) f_{S_i|Y^{i-1}}(s_i|y^{i-1})\,ds_i. \tag{29}$$

Furthermore, it follows from the additive noise model for $y_i|H_1$ that

$$f_{Y_i|Y^{i-1},S_i}(y_i|y^{i-1},s_i) = f_{N_i|Y^{i-1}}(y_i - s_i|y^{i-1}), \tag{30}$$

and it follows from the independence of $\{N_i\}_1^m$ and $\{S_i\}_1^m$, and the independence of the individual noise variables N_i and N_j that (30) reduces to

$$f_{Y_i|Y^{i-1},S_i}(y_i|y^{i-1},s_i) = f_{N_i}(y_i - s_i). \tag{31}$$

Approximation of $f_{N_i}(\cdot)$ by the Taylor series (15), and substitution of (28)–(31) into the preceding expression for $L(y^i)$ yields

$$L(y^i) \simeq L(y^{i-1})\left[\frac{\int\left[f_{N_i}(y_i) - s_i df_{N_i}(y_i)/dy_i + \frac{1}{2}s_i^2 d^2 f_{N_i}(y_i)/dy_i^2\right] f_{S_i|Y^{i-1}}(s_i|y^{i-1})\,ds_i}{f_{N_i}(y_i)}\right]. \tag{32}$$

But (32) reduces to (using (16))

$$L(y^i) \simeq L(y^{i-1})\Big[1 - E\{S_i|y^{i-1}\}dg_i(y_i)/dy_i + \frac{1}{2}E\{(S_i)^2|y^{i-1}\}\big(d^2g_i(y_i)/dy_i^2 + \{dg_i(y_i)/dy_i\}^2\big)\Big], \tag{32'}$$

where $g_i(y_i) \triangleq \log f_{N_i}(y_i)$. Finally the log likelihood ratio $\tau(y^i) = \log L(y^i)$ is, from (32′) (using $\log(1-x) \simeq -x$ since $x \to 0$ as SNR $\to 0$),

$$\tau(y^i) \simeq \tau(y^{i-1}) - \hat{s}_i(y^{i-1})dg_i(y_i)/dy_i + \frac{1}{2}E\{(S_i)^2|y^{i-1}\}\big(d^2g_i(y_i)/dy_i^2 + \{dg_i(y_i)/dy_i\}^2\big), \tag{33}$$

where $\hat{s}_i(y^{i-1}) = E\{S_i|y^{i-1}\}$ is the one-step predicting estimator for S_i and similarly $E\{(S_i)^2|y^{i-1}\}$ is the one-step prediction of $(S_i)^2$. Therefore, $\tau(y^i)$ can be computed recursively if these two predictions can be computed recursively. Equation (33), when approximated by deleting the term involving $E\{(S_i)^2|y^{i-1}\}$, agrees with Sosulin's formula [27, eq. (1.12)], which was derived for a Markov signal process.

In order to relate (33) to the canonical structure (27), we use the locally optimum form (13) for $\hat{s}_i(y^{i-1})$, viz.,

$$\hat{s}_i(y^{i-1}) \simeq \bar{s}_i + \sum_{j=1}^{i-1} K_S(i,j)z_j(y_j), \tag{34}$$

where $z_i(\cdot)$ is defined by (26). Similarly, the predicting estimator of $(S_i)^2$ can be approximated (replacing S_i with $(S_i)^2$ in (34) and deleting the higher order term which $\to 0$ as SNR $\to 0$) by

$$E\{(S_i)^2|y^{i-1}\} \simeq \overline{s_i^2}. \tag{35}$$

Substitution of (34) and (35) into (33), use of (12) and (13) (and use of $1 - \frac{1}{2}\bar{s}_i dg_i(y_i)/dy_i \simeq 1$ as done in deriving (18) from (17)) yields the recursion

$$\tau(y^i) \simeq \tau(y^{i-1}) + \hat{s}_i(y^i)z_i(y_i) + \frac{1}{2}\left[x_i(y_i) - \left(\sigma_{S_i}z_i(y_i)\right)^2\right] \tag{36}$$

where $x_i(\cdot)$ is defined by (26), and $\hat{s}_i(y^i)$ is given by (12). Summation of (36), and use of $\tau(y^0) \triangleq 0$, yields the canonical form (27).

VI. Coherent/Noncoherent Detection of Cyclostationary Signals

If the signal s to be detected is cyclostationary [24] with period, say, p, then the covariance matrix $\boldsymbol{K_S}$ is periodic

(block-Toeplitz)

$$K_S(i+p, j+p) = K_S(i, j). \tag{37}$$

As a result both upper and lower paths in the canonical detector structure in Fig. 2 contain a periodically time-varying transformation (viz., $\hat{s}(\cdot)$) that must be synchronized to the periodicity in the observed process (when the signal is present). This usually requires estimation of the phase (time origin) of the signal, from the observations. When such phase estimation is impractical (as it is in many applications), a noncoherent detector must be employed. The optimum noncoherent detector is simply the optimum detector for the phase-randomized (stationarized) signal (denoted by $\tilde{s}$). Since the locally optimum detector shown in Fig. 2 depends on the signal to be detected through only its mean vector and covariance matrix, then the only difference between the locally optimum coherent and noncoherent detectors is that the signal mean and covariance in the coherent detector are replaced by their time-averaged (stationarized) versions [25] in the noncoherent detector:

$$\bar{\tilde{s}}_i \triangleq \frac{1}{p} \sum_{q=1}^{p} \bar{s}_{i+q},$$

$$K_{\tilde{S}}(i-j) \triangleq \frac{1}{p} \sum_{q=1}^{p} K_S(i+q, j+q). \tag{38}$$

That is, $\bar{s}$ and $\hat{s}(y)$ are replaced with $\bar{\tilde{s}}$ and $\hat{\tilde{s}}(y)$, where $\hat{\tilde{s}}(y)$ is the local-MMSE estimator for the phase-randomized signal $\tilde{s}$. This provides a general framework within which coherent and noncoherent detector structures can be compared (cf. [3]). This relatively simple characterization of the relationship between these two types of structures includes as special cases, and/or provides alternative interpretations of, the results for various specific problems involving random/deterministic amplitude and phase, slow/fast fading channels, narrow-band/broad-band and analog/digital modulation [14]–[16], [29, and refs. therein].

VII. Concluding Remarks

The new structural characterizations shown in Figs. 2(b) and 5(b) provide the basis for extending the estimator–correlator design philosophy, referred to in the Introduction, from continuous-time detectors to discrete-time detectors. This extension of estimator–correlator structures from continuous time to discrete time is, in fact, a generalization since it is valid for non-Gaussian as well as Gaussian noise, a feature which is absent from the structural characterization for continuous-time detectors. The design philosophy is, specifically, to try using practical but suboptimal implementations of estimators (e.g., adaptive estimators) in place of the optimum (or locally optimum) estimators in the estimator–correlator structure. If the resultant detection performance for a trial implementation of an estimator proves to be acceptable in practice (or in simulations), then the approach is successful. This design philosophy has often been successful for continuous-time detection as discussed in the Introduction. The new structural characterizations reveal not only what to correlate the signal estimate with (viz., the nonlinearly transformed observations, $-\nabla g(y)$), but also reveal explicitly what the signal estimator to be approximated with a practical implementation is (cf. Fig. 1). This should be contrasted with the globally optimum characterizations that provide no more than the implicit conditional mean characterization of the signal estimator.

Aside from their fundamental roles in the canonical structures of the locally optimum detectors, the locally optimum estimators (the smoother, filter, and predictor) may be useful in their own right. That is, their relatively simple explicit formulas (11)–(13) may render them useful as suboptimum estimators for applications in which the SNR is not low. The fact that these estimators are recursive for white (non-Gaussian, nonstationary, in general) noise and wide-sense Markov signals is of particular utility. Nevertheless, it can be shown that without some modification, locally optimum estimators will probably perform unnecessarily poorly when the SNR is not low. This is easily demonstrated for Gaussian noise, using the results in Section II-B. Moreover, the results in Section II-B suggest the following *modified version* of the locally optimum estimator (11)

$$\hat{s}(y) = \bar{s} + K_S[K_S + K_N]^{-1} K_N[-\nabla g(y - \bar{s})],$$

which is *optimum* for Gaussian noise. Since the modification is negligible for low SNR, then this modified version is still locally optimum, regardless of the noise distribution.

There are signal detection problems (other than the additive noise problems addressed herein) for which the optimum detector structure is quite simple, except for the computation of the MMSE estimator (e.g., jump process observations [4], [8]–[11]). For such problems, a possibly adequate suboptimum detector can be obtained by use of the local-MMSE estimator in place of the global-MMSE estimator. For example, it has been shown [5] that an adequate suboptimum detector for Poisson observations is obtained by replacement of the MMSE nonlinear estimator with the linear MMSE estimator (cf. [26]). Therefore, a possibly fruitful topic for future research is the study of local-MMSE estimators for observations other than signal in additive noise.

The random signal in additive random noise problem addressed in this paper includes the signal (both random and nonrandom) in multiplicative and additive random noise problem simply by redefining the signal S to be $S = [M]S_0$, where S_0 is the actual signal, and $[M]$ is a diagonal matrix representing the multiplicative noise vector M. (More generally, $[M]$ need not be diagonal.) The net result is that the matrix K_S in the canonical estimators and detectors decomposes into $K_S = E\{[M]K_{S_0}[M]^T\}$, which further reduces for diagonal $[M]$ to $K_S = K_M \otimes K_{S_0}$, where $\otimes$ denotes Schur matrix product, $K_S(i, j) = K_M(i, j)K_{S_0}(i, j)$ [28, p. 646], for which $K_M(i, j)$ is the correlation of the ith and jth diagonal elements of $[M]$. In addition, the mean vector $\bar{s}$ in the canonical estimators and detectors decomposes into $\bar{s} = E\{[M]\}\bar{s}_0$.

References

[1] R. Price and P. E. Green, "A communication technique for multipath channels," *Proc. IEEE*, vol. 46, pp. 555–570, Mar. 1958.

[2] R. Price, "Optimum detection of random signals in noise, with application to scatter-multipath communication—I," *IEEE Trans. Inform. Theory*, vol. IT-2, pp. 125–135, Dec. 1956.

[3] W. A. Gardner and D. C. Bukofzer, "Performance evaluation of suboptimum estimator-correlator detectors for Gaussian signals with low energy coherence," Signal and Image Processing Lab. Tech. Rep. no. SIPL-80-15, Univ. of California, Davis, CA, Dec. 1980.

[4] R. Boel, P. P. Varaiya, and E. Wong, "Martingales on jump processes," *SIAM J. Contr.*, vol. 13, pp. 999–1061, 1975.

[5] M. H. A. Davis and E. Andreadakis, "Exact and approximate filtering in signal detection: An example," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 768–772, Nov. 1977.

[6] T. Kailath, "The innovations approach to detection and estimation theory," *Proc. IEEE*, vol. 58, pp. 680–695, 1970.

[7] ——, "The structure of Radon–Nikodym derivatives with respect to Wiener and related measures," *Ann. Math. Statist.*, vol. 42, pp. 1054–1067, 1971.

[8] D. L. Snyder, "Filtering and detection for doubly stochastic Poisson processes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 91–102, 1972.

[9] I. Rubin, "Regular point processes and their detection," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 547–557, 1972.

[10] P. M. Bremaud, "A Martingale approach to point processes," Report ERL-M345, College of Eng., Univ. of CA, Berkeley, Aug. 1972.

[11] A. Segal, "A Martingale approach to modeling, estimation, and detection of jump processes," Report no. 7050-21, Information Systems Lab., Stanford Univ., Stanford, CA, Aug. 1973.

[12] L. L. Scharf and L. W. Nolte, "Likelihood ratios for sequential hypothesis testing on Markov sequences," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 101–109, 1977.

[13] S. C. Schwartz, "The Estimator–Correlator for discrete-time problems," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 93–100, 1977.

[14] D. Middleton, "Canonically optimum threshold detection," *IEEE Trans. Inform. Theory*, vol. IT-12, no. 2, pp. 230–243, 1966.

[15] A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment—Part I: Coherent detection," *IEEE Trans. Commun.*, vol. COM-25, pp. 910–923, Sept. 1977.

[16] ——, "Optimum reception in an impulsive interference environment—Part II: Incoherent reception," *IEEE Trans. Commun.*, vol. COM-25, pp. 924–934, Sept. 1977.

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

[18] J. W. Modestino and A. Y. Ningo, "Detection of weak signals in narrow-band non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 592–600, Sept. 1979.

[19] J. Capon, "On asymptotic relative efficiency of locally optimum detectors," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67–72, Apr. 1961.

[20] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 241–250, Mar. 1972.

[21] A. F. Kushner and B. R. Levin, "Optimal ranking algorithms for the detection of signals in noise," *Probl. Peredach. Inform.*, vol. 4, 1968.

[22] A. K. Ribin, "Classification of weak signals in non-Gaussian noise," *Eng. Cybern.* (USSR), vol. 10, pp. 901–909, 1972.

[23] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III.* New York: Wiley, 1971.

[24] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 4–14, Jan. 1975.

[25] W. A. Gardner, "Stationarizable random processes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 8–22, Jan. 1978.

[26] ——, "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.

[27] Y. G. Sosulin, "Estimation-correlation methods for detection of weak signals in noise and mathematical-heuristic synthesis," *Eng. Cybern.* (USSR), vol. 9, pp. 909–918, 1971.

[28] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[29] N. H. Lu and B. A. Eisenstein, "Detection of weak signals in non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 755–771, Nov. 1981.