

Neyman–Pearson Detection of Gauss–Markov Signals in Noise: Closed-Form Error Exponent and Properties

Youngchul Sung, *Member, IEEE*, Lang Tong, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

Abstract—The performance of Neyman–Pearson detection of correlated random signals using noisy observations is considered. Using the large deviations principle, the performance is analyzed via the error exponent for the miss probability with a fixed false-alarm probability. Using the state-space structure of the signal and observation model, a closed-form expression for the error exponent is derived using the innovations approach, and the connection between the asymptotic behavior of the optimal detector and that of the Kalman filter is established. The properties of the error exponent are investigated for the scalar case. It is shown that the error exponent has distinct characteristics with respect to correlation strength: for signal-to-noise ratio (SNR) ≥ 1 , the error exponent is monotonically decreasing as the correlation becomes strong whereas for SNR < 1 there is an optimal correlation that maximizes the error exponent for a given SNR.

Index Terms—Autoregressive process, correlated signal, error exponent, Gauss–Markov model, Neyman–Pearson detection.

I. INTRODUCTION

IN this paper, we consider the detection of correlated random signals using noisy observations y_i under the Neyman–Pearson formulation. The null and alternative hypotheses are given by

$$\begin{aligned} H_0 : y_i &= w_i, & i &= 1, 2, \dots, n \\ H_1 : y_i &= s_i + w_i, & i &= 1, 2, \dots, n \end{aligned} \quad (1)$$

where $\{w_i\}$ is independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$ noise with a known variance σ^2 , and $\{s_i\}$ is the stochastic signal process correlated in time. We assume that $\{s_i\}$ is a Gauss–Markov process following a state-space model. An example of an application in which this type of problem arises

Manuscript received November 18, 2004; revised December 30, 2005. This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The work of H. V. Poor was also supported in part by the Office of Naval Research under Grant N00014-03-1-0102. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September, 2005.

Y. Sung is with Qualcomm Inc., San Diego, CA 92121, USA (e-mail: ysung@qualcomm.com).

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 (e-mail: ltong@ece.cornell.edu).

H. V. Poor is with Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2006.871599

is the detection of stochastic signals in large sensor networks, where it is reasonable to assume that signal samples taken at closely spaced locations are correlated, while the measurement noise is independent from sensor to sensor. In this paper, we are interested in the performance of the Neyman–Pearson detector for the hypotheses (1) with a fixed level (i.e., upper-bound constraint on the false-alarm probability) when the sample size n is large.

In many cases, the miss probability P_M of the Neyman–Pearson detector with a fixed level decays exponentially as the sample size increases, and the error exponent is defined as the rate of exponential decay, i.e.,

$$K \triangleq \lim_{n \rightarrow \infty} -\frac{1}{n} \log P_M \quad (2)$$

under the given false-alarm constraint. The error exponent is an important parameter since it gives an estimate of the number of samples required for a given detector performance; faster decay rate implies that fewer samples are needed for a given miss probability, or that better performance can be obtained with a given number of samples. Hence, the error exponent is a good performance index for detectors in the large sample regime. For the case of i.i.d. samples, where each sample is drawn independently from the common null probability density p_0 or alternative density p_1 , the error exponent under the fixed false-alarm constraint is given by the Kullback–Leibler information $D(p_0||p_1)$ between the two densities p_0 and p_1 (see [29]). For more general cases, the error exponent is given by the asymptotic Kullback–Leibler rate defined as the almost-sure limit of

$$\frac{1}{n} \log \frac{p_{0,n}}{p_{1,n}}(y_1, \dots, y_n), \quad \text{as } n \rightarrow \infty \quad (3)$$

under $p_{0,n}$, where $p_{0,n}$ and $p_{1,n}$ are the null and alternative joint densities of y_1, \dots, y_n , respectively, assuming that the limit exists [30]–[34].¹ However, the closed-form calculation of (3) is available only for restricted cases. One such example is the discrimination between two autoregressive (AR) signals with distinct parameters under the two hypotheses [34], [35]. In this case, the joint density $p_{j,n}$ is easily decomposed using the Markov property under each hypothesis, and the calculation of the rate is straightforward. However, for the problem of (1) this approach is not available since the observation samples under the alternative hypothesis do not possess the Markov property due to the additive noise, even if the signal itself is Markovian; i.e., the alternative is a hidden Markov model.

¹Ergodic cases are examples for which this limit exists.

A. Summary of Results

Our approach to this problem is to exploit the state-space model. The state-space approach in detection is well established in calculation of the log-likelihood ratio (LLR) for correlated signals [6], [7]. With the state-space model, the LLR is expressed through the innovations representation [11] and the innovations are easily obtained by the Kalman filter. The key idea for the closed-form calculation of the error exponent for the hidden Markov model is based on the properties of innovations. Since the innovations process is independent from time to time, the joint density under H_1 is given by the product of marginal densities of the innovations, and the LLR is given by a function of the sum of squares of the innovations; this functional form facilitates the closed-form calculation of (3).

By applying this state-space approach, we derive a closed-form expression for the error exponent K for the miss probability of the Neyman–Pearson detector for (1) of fixed false-alarm probability $P_F = \alpha$.

We next investigate the properties of the error exponent using the obtained closed-form expression. We explore the asymptotic relationship between the innovations approach and the spectrum of the observation. We show that the error exponent K is a function of the signal-to-noise ratio (SNR) and the correlation, and has different behavior with respect to (w.r.t.) the correlation strength depending on the SNR. We show a sharp phase transition at SNR = 1: at high SNR, K is monotonically decreasing as a function of the correlation, while at low SNR, on the other hand, there exists an optimal correlation value that yields the maximal K .

We also make a connection between the asymptotic behavior of the Kalman filter and that of the Neyman–Pearson detector. It is shown that the error exponent is determined by the asymptotic (or steady-state) variances of the innovations under H_0 and H_1 together with the noise variance.

B. Related Work

The detection of Gauss–Markov processes in Gaussian noise is a classical problem. See [5] and references therein. Our work focuses on the performance analysis as measured by the error exponent, and relies on the connection between the likelihood ratio and the innovations process as described by Schweppe [6]. In addition to the calculation of the LLR, the state-space approach has also been used in the performance analysis in this detection problem. Exploiting the state-space model, Schweppe obtained a differential equation for the Bhattacharyya distance between two Gaussian processes [8]–[10], which gives an upper bound on the average error probability under a Bayesian formulation.

There is an extensive literature on the large deviations approach to the analysis of the detection of Gauss–Markov processes [19]–[28]. Many of these results rely on the extension of Cramer’s theorem by Gärtner and Ellis [15]–[18] and the properties of the asymptotic eigenvalue distributions of Toeplitz matrices [12], [13]. To find the rate function, however, this approach usually requires an optimization that requires non-trivial numerical methods except in some simple cases, and the rate is given as an integral of the spectrum of the observation

process; closed-form expressions are difficult to obtain except for the case of a noiseless AR process in discrete time and its continuous-time counterpart, the Ornstein–Uhlenbeck process [19]–[27]. In addition, most results have been obtained for a fixed threshold for the normalized LLR test, which results in expressions for the rate as a function of the threshold. For ergodic cases, however, the normalized LLR converges to a constant under the null hypothesis and the false-alarm probability also decays exponentially for a fixed threshold. Hence, a detector with a fixed threshold is not optimal in the Neyman–Pearson sense since it does not use the false-alarm constraint fully; i.e., the optimal threshold is a function of sample size.

C. Notation and Organization

We will make use of standard notational conventions. Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a scalar z , z^* denotes the complex conjugate. For a matrix \mathbf{A} , \mathbf{A}^T and \mathbf{A}^H indicate the transpose and Hermitian transpose, respectively. $\det(\mathbf{A})$ and $\text{tr}(\mathbf{A})$ denote the determinant and trace of \mathbf{A} , respectively. $\mathbf{A}(l, m)$ denotes the element of the l th row and m th column, and $\{\lambda_k(\mathbf{A})\}$ denotes the set of all eigenvalues of \mathbf{A} . We reserve \mathbf{I}_m for the identity matrix of size m (the subscript is included only when necessary). For a sequence of random vectors \mathbf{x}_n , $\mathbb{E}_j\{\mathbf{x}_n\}$ is the expectation of \mathbf{x}_n under probability density $p_{j,n}$, $j = 0, 1$. The notation $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that \mathbf{x} has the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

The paper is organized as follows. The data model is described in Section II. In Section III, the closed-form error exponent is obtained via the innovations approach representation. The properties of the error exponent are investigated in Section IV, and the extension to the vector case is provided in Section V. Simulation results are presented to demonstrate the predicted behavior in Section VI, followed by the conclusion in Section VII.

II. DATA MODEL

For the purposes of exposition, we will focus primarily on the case in which the signal is generated by a scalar time-invariant state-space model. The more general vector case will be considered below. In particular, we assume that the signal process $\{s_i\}$ has a time-invariant state-space structure

$$\begin{aligned} s_{i+1} &= as_i + u_i, & i &= 1, \dots, n \\ s_1 &\sim \mathcal{N}(0, \Pi_0) \\ u_i &\stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q), & Q &= \Pi_0(1 - a^2) \end{aligned} \quad (4)$$

where a and Π_0 are known scalars with $0 \leq a \leq 1$ and $\Pi_0 \geq 0$. We assume that the process noise $\{u_i\}$ is independent of the measurement noise $\{w_i\}$ and the initial state s_1 is independent of u_i for all i . Notice that the signal sequence $\{s_i\}$ forms a stationary process for this choice of Q . Due to this stationarity, the signal variance is Π_0 for all i , and the SNR Γ for the observations is thus given by

$$\Gamma = \frac{\Pi_0}{\sigma^2}. \quad (5)$$

Notice that the value of a determines the amount of correlation between signal samples. For an i.i.d. signal we have $a = 0$ and all the signal power results from the process noise $\{u_i\}$. When the signal is perfectly correlated on the other hand, $a = 1$ and the signal depends only on the realization of the initial state s_1 . The autocovariance function $r_s(\cdot)$ of the signal process $\{s_i\}$ is given by

$$r_s(i-j) \triangleq \mathbb{E}\{s_i s_j\} = \Pi_0 a^{|i-j|}. \quad (6)$$

As seen in (1), the observation y_i under the alternative hypothesis is given by a sum of signal sample s_i and independent noise w_i . Thus, the observation sequence $\{y_i\}$ under H_1 is not a Markov process due to the presence of the additive noise even if the signal is Markovian. Let $r_y^{(j)}(\cdot)$ denote the autocovariance function of the observation process $\{y_i\}$ under H_j , i.e.,

$$r_y^{(j)}(m-n) = \mathbb{E}_j\{y_m y_n\} \quad (7)$$

and let $S_y^{(j)}(\omega)$ be the spectrum of the observation process under H_j , i.e.,

$$S_y^{(j)}(\omega) = \sum_{k=-\infty}^{\infty} r_y^{(j)}(k) e^{-jk\omega}, \quad -\pi \leq \omega \leq \pi. \quad (8)$$

Then, the spectra of the observation process under H_0 and H_1 are given by

$$S_y^{(0)}(\omega) = \sigma^2, \quad S_y^{(1)}(\omega) = \sigma^2 + S_s(\omega), \quad -\pi \leq \omega \leq \pi \quad (9)$$

where the signal spectrum under the state-space model is given by the Poisson kernel

$$S_s(\omega) = \frac{\Pi_0(1-a^2)}{1-2a\cos\omega+a^2}, \quad 0 \leq a < 1. \quad (10)$$

III. ERROR EXPONENT FOR GAUSS-MARKOV SIGNAL IN NOISE

In this section, we derive the error exponent of the Neyman-Pearson detector with a fixed level $\alpha \in (0, 1)$ for the Gauss-Markov signal described by (4) embedded in noisy observations.

A general approach to the error exponent of Neyman-Pearson detection of Gaussian signals can be framed in the spectral domain. It is well known that the Kullback-Leibler information between two zero-mean Gaussian distributions $p_0 = \mathcal{N}(0, \sigma_0^2)$ and $p_1 = \mathcal{N}(0, \sigma_1^2)$ is given by

$$D(p_0||p_1) = \frac{1}{2} \log \frac{\sigma_1^2}{\sigma_0^2} + \frac{1}{2} \frac{\sigma_0^2}{\sigma_1^2} - \frac{1}{2}. \quad (11)$$

As noted earlier, this quantity gives the error exponent in the case of an i.i.d. Gaussian signal. In more general cases with correlated Gaussian signals, the error exponent can similarly be obtained using the asymptotic properties of covariance matrices. Let \mathbf{y}_n be the random vector of observation samples y_i defined as

$$\mathbf{y}_n \triangleq [y_1, y_2, \dots, y_n]^T. \quad (12)$$

For two distributions

$$p_{0,n}(\mathbf{y}_n) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{0,n}) \quad \text{and} \quad p_{1,n}(\mathbf{y}_n) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{1,n})$$

the error exponent is given by the almost-sure limit of the Kullback-Leibler rate

$$\begin{aligned} & \frac{1}{n} \log \frac{p_{0,n}(\mathbf{y}_n)}{p_{1,n}(\mathbf{y}_n)} \\ &= \frac{1}{n} \left(\frac{1}{2} \log \left(\frac{\det(\boldsymbol{\Sigma}_{1,n})}{\det(\boldsymbol{\Sigma}_{0,n})} \right) + \frac{1}{2} \mathbf{y}_n^T (\boldsymbol{\Sigma}_{1,n}^{-1} - \boldsymbol{\Sigma}_{0,n}^{-1}) \mathbf{y}_n \right) \end{aligned} \quad (13)$$

under $p_{0,n}$ [30]–[34]. Using the asymptotic distribution of the eigenvalues of a Toeplitz matrix [12], [13], we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\det(\boldsymbol{\Sigma}_{j,n})) = \frac{1}{2\pi} \int_0^{2\pi} \log S_y^{(j)}(\omega) d\omega, \quad j = 0, 1 \quad (14)$$

where $S_y^{(j)}(\omega)$ is the spectrum of $\{y_i\}$ which is assumed to have finite lower and upper bounds under distribution $\mathbf{y}_n \sim p_{j,n}$. The limiting behavior of $n^{-1} \mathbf{y}_n^T \boldsymbol{\Sigma}_{j,n}^{-1} \mathbf{y}_n$ is also known and is given by (assuming that the true distribution of $\{y_i\}$ is $p_{0,n}$)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{y}_n^T \boldsymbol{\Sigma}_{1,n}^{-1} \mathbf{y}_n = \frac{1}{2\pi} \int_0^{2\pi} \frac{S_y^{(0)}(\omega)}{S_y^{(1)}(\omega)} d\omega \quad (15)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{y}_n^T \boldsymbol{\Sigma}_{0,n}^{-1} \mathbf{y}_n = 1 \quad (16)$$

where the limit is in the almost-sure sense convergence under H_0 , provided that $S_y^{(0)}(\omega)$ and $S_y^{(1)}(\omega)$ are continuous and strictly positive. (See [14, Lemmas 1 and 2] and [4, Propositions 10.8.2 and 10.8.3].) Combining (13)–(16), the error exponent for two zero-mean stationary Gaussian processes is thus given by

$$K = \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{p_{0,n}(\mathbf{y}_n)}{p_{1,n}(\mathbf{y}_n)} \quad \text{a.s. } [p_{0,n}] \quad (17)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{1}{2} \log \frac{S_y^{(1)}(\omega)}{S_y^{(0)}(\omega)} + \frac{S_y^{(0)}(\omega)}{2S_y^{(1)}(\omega)} - \frac{1}{2} \right) d\omega \quad (18)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} D(\mathcal{N}(0, S_y^{(0)}(\omega)) || \mathcal{N}(0, S_y^{(1)}(\omega))) d\omega. \quad (19)$$

Intuitively, the error exponent (19) can be explained from (11) using the frequency binning argument used to obtain the channel capacity of Gaussian channel with colored noise from that of independent parallel Gaussian channels [3].

The spectral form (19) of the error exponent is valid for a wide class of stationary Gaussian processes including the autoregressive moving average (ARMA) processes and the hidden Markov model (1)–(4). For the detection (1) under the scalar state-space model (4), we have

$$\boldsymbol{\Sigma}_{0,n} = \mathbf{I}, \quad \boldsymbol{\Sigma}_{1,n} = \boldsymbol{\Sigma}_{s,n} + \mathbf{I} \quad (20)$$

where $\boldsymbol{\Sigma}_{s,n}(l, m) = \Pi_0 a^{|l-m|}$, $l, m = 1, \dots, n$, and two spectra under H_0 and H_1 are given by (9). However, it is not straightforward to obtain a closed-form expression for (19) except in some special cases, e.g., when both of the

two distributions of $\{y_i\}$ under H_0 and H_1 have the Markov property [35].

In the remainder of the paper, we focus on the derivation of a closed-form expression for the error exponent K of the miss probability for (1)–(4) by exploiting the state-space structure under the alternative hypothesis. We do so by making a connection with Kalman filtering [11]. Our expressions will allow us to investigate the properties of the error exponent.

A. Closed-Form Error Exponent via Innovations Approach

Theorem 1 (Error Exponent): For the Neyman–Pearson detector for the hypotheses (1) with level $\alpha \in (0, 1)$ (i.e., $P_F \leq \alpha$) and $0 \leq a \leq 1$, the error exponent of the miss probability is given by

$$K = \frac{1}{2} \log \frac{R_e}{\sigma^2} + \frac{1}{2} \frac{\tilde{R}_e}{R_e} - \frac{1}{2} \quad (21)$$

independently of the value of α , where R_e and \tilde{R}_e are the steady-state variances of the innovations process of $\{y_i\}$ calculated under H_1 and H_0 , respectively. Specifically, R_e and \tilde{R}_e are given by

$$R_e = P + \sigma^2 \quad (22)$$

and

$$\tilde{R}_e = \sigma^2 \left(1 + \frac{a^2 P^2}{P^2 + 2\sigma^2 P + (1 - a^2)\sigma^4} \right) \quad (23)$$

where

$$P = \frac{1}{2} \sqrt{[\sigma^2(1 - a^2) - Q]^2 + 4\sigma^2 Q} - \frac{1}{2} \sigma^2(1 - a^2) + \frac{Q}{2}. \quad (24)$$

Here, P is the steady-state error variance of the minimum mean-square error (MMSE) estimator for the signal s_i under the model H_1 . Note that the error exponent (21) is thus a closed form of (19) for the state-space model.

Proof: See the Appendix .

Theorem 1 follows from the fact that the almost-sure limit (3) of the normalized log-likelihood ratio under H_0 is the error exponent for general ergodic cases [31]–[34]. To make the closed-form calculation of the error exponent tractable for the hidden Markov structure of $\{y_i\}$, we express the LLR through the innovations representation [6]; the LLR is given by a function of the sum of squares of the innovations on which the strong law of large numbers (SLLN) is applied. The calculated innovations are true in the sense that they form an independent sequence only under H_1 , i.e., when the signal actually comes from the state-space model. It is worth noting that \tilde{R}_e is the steady-state variance of the “innovations” calculated as if the observations resulted from the alternative, but are actually from the null hypothesis. In this case, the “innovation” sequence becomes the output of a recursive (whitening) filter driven by an i.i.d. process $\{y_i\}$ since the Kalman filter converges to the recursive Wiener filter for time-invariant stable systems [2].

The relationship between the spectral-domain approach and the innovations approach is explained by the canonical spectral factorization, which is well established for the state-space model. The asymptotic variance of the innovations sequence is the key parameter in both cases. The relationship between the asymptotic performance of the Neyman–Pearson detector and that of the Kalman filter is evident in (21) for the state-space model. In both cases, the innovations process plays a critical role in characterizing the performance, and the asymptotic variance of the innovation is sufficient for the calculation of the error exponent for the Neyman–Pearson detector and the steady-state error variance for the Kalman filter.

IV. PROPERTIES OF ERROR EXPONENT

In this section, we investigate the properties of the error exponent derived in the preceding section. We particularly examine the large sample error behavior with respect to the correlation strength and SNR. We show that the intensity of the additive noise significantly changes the error behavior with respect to the correlation strength, and the error exponent has a distinct phase transition in behavior with respect to the correlation strength depending on SNR.

Theorem 2 (K Versus Correlation): The error exponent K is a continuous function of the correlation coefficient a ($0 \leq a \leq 1$) for a given SNR ≥ 0 . The error exponent as a function of correlation strength is characterized by the following.

- i) For SNR $\Gamma \geq 1$, K is monotonically decreasing as the correlation strength increases (i.e., $a \uparrow$ 1).
- ii) For SNR $\Gamma < 1$, there exists a nonzero value a^* of the correlation coefficient that achieves the maximal K , and a^* is given by the solution of the following equation:

$$[1 + a^2 + \Gamma(1 - a^2)]^2 - 2 \left(r_e + \frac{a^4}{r_e} \right) = 0 \quad (25)$$

where $r_e = R_e/\sigma^2$. Furthermore, a^* converges to one as Γ goes to zero.

Proof: See the Appendix.

We first note that Theorem 2 shows that an i.i.d. signal gives the best error performance for a given SNR ≥ 1 with the maximal error exponent being $D(\mathcal{N}(0, \sigma^2) || \mathcal{N}(0, \Pi_0 + \sigma^2))$. (In this case, Theorem 1 reduces to Stein’s lemma.) The intuition behind this result is that the signal component in the observation is strong at high SNR, and the innovations (the new information) provide more benefit to the detector than the noise averaging effect present for correlated observations. That is, simple radiometry provides sufficient detection power when the signal level is above that of the noise. Fig. 1 shows the error exponent as a function of the correlation coefficient a for SNR $\Gamma = 10$ dB. The monotonicity of the error exponent is clearly seen; moreover, we see that the amount of decrease becomes larger as a increases. Notice also that the amount of performance degradation from the i.i.d. case is not severe for weak correlation and the error exponent decreases suddenly near $a = 1$ and eventually becomes zero at $a = 1$. (It is easy to show that the miss probability decays with $\Theta(\frac{1}{\sqrt{n}})$ for any SNR at $a = 1$.)

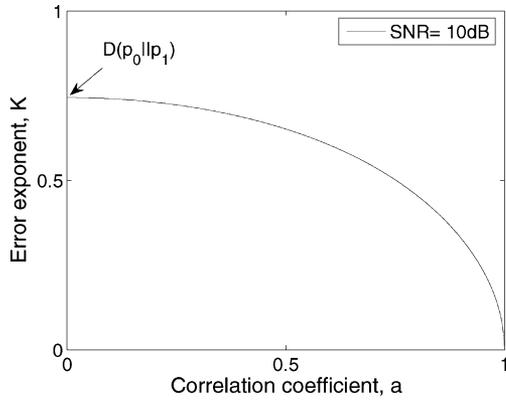


Fig. 1. K versus correlation coefficient a (SNR =) 10 dB: $p_0 = \mathcal{N}(0, \sigma^2)$, $p_1 = \mathcal{N}(0, \Pi_0 + \sigma^2)$.

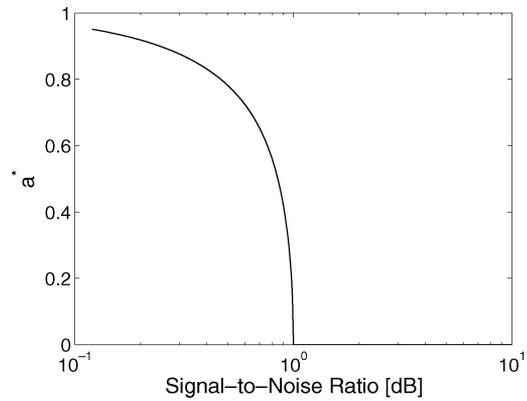


Fig. 3. Optimum correlation strength versus SNR.

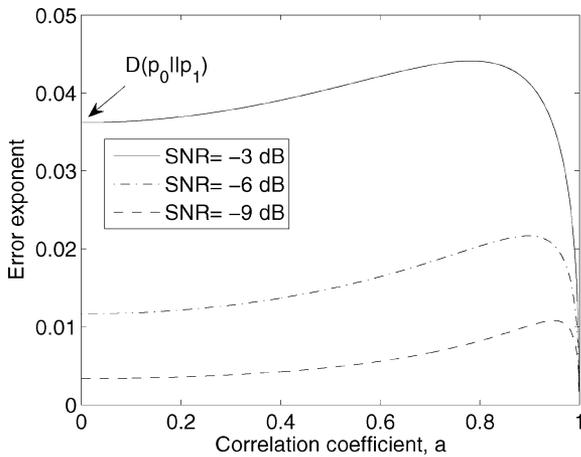


Fig. 2. K versus correlation coefficient a (SNR = -3, -6, -9 dB).

In contrast, the error exponent does not decrease monotonically in a for SNR < 1, and there exists an optimal correlation as shown in Fig. 2. It is seen that the i.i.d. case no longer gives the error performance for a fixed SNR. The error exponent initially increases as a increases, and then decreases to zero as a approaches one. As the SNR further decreases (see the cases of -6 and -9 dB) the error exponent decreases for a fixed correlation strength, and the value of a achieving the maximal error exponent is shifted closer to one. At low SNR, the noise in the observation dominates. So, intuitively, making the signal more correlated provides greater benefit of noise averaging. The lower the SNR, the stronger we would like the correlation to be in order to compensate for the dominant noise power, as shown in Fig. 2. However, excessive correlation in the signal does not provide new information by observation, and the error exponent ultimately converges to zero as a approaches one. Notice that the ratio of the error exponent for the optimal correlation to that for the i.i.d. case becomes large as SNR decreases. Hence, the improvement due to optimal correlation can be large for low SNR cases. Fig. 3 shows the value of a that maximizes the error exponent as a function of SNR. As shown in the figure, unit SNR is a transition point between two different behavioral regimes of the error exponent with respect to correlation strength, and the transition is very sharp; the optimal correlation strength a approaches one rapidly once SNR becomes smaller than one.

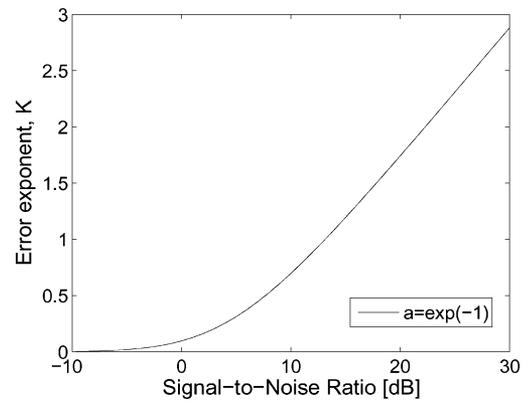


Fig. 4. K versus SNR ($a = e^{-1}$).

The behavior of the error exponent with respect to SNR is given by the following theorem.

Theorem 3 (K Versus SNR): The error exponent K is monotonically increasing as SNR increases for a given correlation coefficient $0 \leq a < 1$. Moreover, at high SNR, the error exponent K increases linearly with respect to $\frac{1}{2} \log[1 + \text{SNR}(1 - a^2)]$.

Proof: See the Appendix.

The detrimental effect of correlation at high SNR is clear.² The performance degradation due to correlation is equivalent to the SNR decreasing by factor $(1 - a^2)$. The $\log(1 + \text{SNR})$ increase of K w.r.t. SNR is analogous to similar error-rate behavior arising in diversity combining of versions of a communications signal arriving over independent Rayleigh-faded paths in additive noise, where the error probability is given by $P_e \sim (1 + \text{SNR})^{-L}$ and L is the number of independent multipaths. In both cases, the signal component is random. The log SNR behavior of the optimal Neyman-Pearson detector for stochastic signals applies to general correlations as well with a modified definition of SNR. Comparing with the detection of a deterministic signal in noise, where the error exponent is proportional to SNR, the increase of error exponent w.r.t. SNR is much slower for the case of a stochastic signal in noise. Fig. 4 shows the error exponent with respect to SNR for a given correlation strength. The log SNR behavior is evident at high SNR.

²Interestingly, the error exponent at high SNR has the same expression as the capacity of the Gaussian channel.

V. EXTENSION TO THE VECTOR CASE

In order to treat general cases in which the signal is a higher order AR process or the signal is determined by a linear combination of several underlying phenomena, we now consider a vector state-space model, and extend the results of the previous sections to this model. The hypotheses for the vector case are given by

$$\begin{aligned} H_0 : y_i &= w_i, & i &= 1, 2, \dots, n \\ H_1 : y_i &= \mathbf{h}^T \mathbf{s}_i + w_i, & i &= 1, 2, \dots, n \end{aligned} \quad (26)$$

where \mathbf{h} is a known vector and $\mathbf{s}_i \triangleq [s_{1i}, s_{2i}, \dots, s_{mi}]^T$ is the state of an m -dimensional process at time i following the state-space model

$$\begin{aligned} \mathbf{s}_{i+1} &= \mathbf{A}\mathbf{s}_i + \mathbf{B}\mathbf{u}_i \\ \mathbf{s}_1 &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Pi}_0) \\ \mathbf{u}_i &\stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \mathbf{Q} \geq 0. \end{aligned} \quad (27)$$

We assume that the feedback and input matrices, \mathbf{A} and \mathbf{B} , are known with $|\lambda_k(\mathbf{A})| < 1$ for all k , and the process noise $\{\mathbf{u}_i\}$ independent of the measurement noise $\{w_i\}$. We also assume that the initial state \mathbf{s}_1 is independent of \mathbf{u}_i for all i , and the initial covariance $\mathbf{\Pi}_0$ satisfies the following Lyapunov equation:

$$\mathbf{\Pi}_0 = \mathbf{A}\mathbf{\Pi}_0\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T. \quad (28)$$

Thus, the signal sequence $\{\mathbf{s}_i\}$ forms a stationary vector process. In this case the SNR is defined similarly to (5) as $\frac{\mathbf{h}^T \mathbf{\Pi}_0 \mathbf{h}}{\sigma^2}$. The autocovariance of the observation sequence $\{y_i\}$ is given by

$$r_y(i, j) = \mathbb{E}\{y_i y_j\} = \begin{cases} \sigma^2 \delta_{ij}, & \text{under } H_0 \\ \mathbf{h}^T \mathbf{A}^{|i-j|} \mathbf{\Pi}_0 \mathbf{h} + \sigma^2 \delta_{ij}, & \text{under } H_1 \end{cases} \quad (29)$$

where δ_{ij} is the Kronecker delta. Thus, the covariance matrix of the observation under H_1 is a symmetric Toeplitz matrix with $\mathbf{h}^T \mathbf{A}^l \mathbf{\Pi}_0 \mathbf{h}$ as the l th off-diagonal entry ($l \neq 1$). Since $|\lambda_k(\mathbf{A})| < 1$ for all k , $c_l \triangleq \mathbf{h}^T \mathbf{A}^l \mathbf{\Pi}_0 \mathbf{h}$ is an absolutely summable sequence and the eigenvalues of the covariance matrix of \mathbf{y}_n is bounded both from below and from above.

Theorem 4 (Error Exponent): For the Neyman–Pearson detector for the hypotheses (26) and (27) with level $\alpha \in (0, 1)$ (i.e., $P_F \leq \alpha$) and $|\lambda_k(\mathbf{A})| < 1$ for all k , the error exponent of the miss probability is given by (21) independently of the value of α . The steady-state variances of the innovation process R_e and \tilde{R}_e calculated under H_1 and H_0 , respectively, are given by

$$R_e = \sigma^2 + \mathbf{h}^T \mathbf{P} \mathbf{h} \quad (30)$$

where \mathbf{P} is the unique stabilizing solution of the discrete-time algebraic Riccati equation

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T - \frac{\mathbf{A}\mathbf{P}\mathbf{h}\mathbf{h}^T\mathbf{P}\mathbf{A}^T}{\mathbf{h}^T\mathbf{P}\mathbf{h} + \sigma^2} \quad (31)$$

and

$$\tilde{R}_e = \sigma^2(1 + \mathbf{h}^T \tilde{\mathbf{P}} \mathbf{h}) \quad (32)$$

where $\tilde{\mathbf{P}}$ is the unique positive-semidefinite solution of the following Lyapunov equation:

$$\tilde{\mathbf{P}} = (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) \tilde{\mathbf{P}} (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^T + \mathbf{K}_p \mathbf{K}_p^T \quad (33)$$

and $\mathbf{K}_p = \mathbf{A}\mathbf{P}\mathbf{h}R_e^{-1}$.

In spectral form, K is given by (19), where

$$S_y^{(0)}(\omega) = \sigma^2 (-\pi \leq \omega \leq \pi)$$

and $S_y^{(1)}(\omega)$ is given by

$$S_y^{(1)}(\omega) = [\mathbf{h}^T (e^{j\omega} \mathbf{I} - \mathbf{A})^{-1} \quad 1] \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} (e^{-j\omega} \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{h} \\ 1 \end{bmatrix} \quad (34)$$

Proof: See the Appendix.

For this vector model, simple results describing the properties of the error exponent are not tractable since the relevant expressions depend on the multiple eigenvalues of the matrix \mathbf{A} . However, (21), (31), and (33) provide closed-form expressions for the error exponent which can easily be explored numerically.

VI. SIMULATION RESULTS

To verify the behavior of the miss probability predicted by our asymptotic analysis, in this section we provide some simulation results. We consider the scalar model (4), for SNR of 10 and -3 dB, and for several correlation strengths. The probability of false alarm is set at 0.1% for all cases we consider.

Fig. 5 shows the simulated miss probability as a function of the number of samples for 10-dB SNR. It is seen, as predicted by our analysis, that the i.i.d. case ($a = 0$) has the largest slope for error decay, and the slope is monotonically decreasing as a increases to one. Notice that the error performance for the same number of observations is significantly different for different correlation strengths for the same SNR, and the performance for weak correlation is not much different from the i.i.d. case, as predicted by Fig. 1. It is also seen that the miss probability for the perfectly correlated case ($a = 1$) is not exponentially decaying, again confirming our analysis.

The simulated error performance for SNR of -3 dB is shown in Fig. 6. It is seen that the asymptotic slope of $\log P_M$ increases as a increases from zero as predicted by Theorem 2, and reaches a maximum with a sudden decrease after the maximum. Notice that the error curve is still not a straight line for the low SNR case due to the $o(n)$ term in the exponent of the error probability. Since the error exponent increases only with \log SNR, the required number of observations for -3 -dB SNR is much larger than for 10-dB SNR for the same miss probability. It is clearly seen that P_M is still larger than 10^{-2} for 200 samples whereas it is 10^{-4} with 20 samples for the 10-dB SNR case.

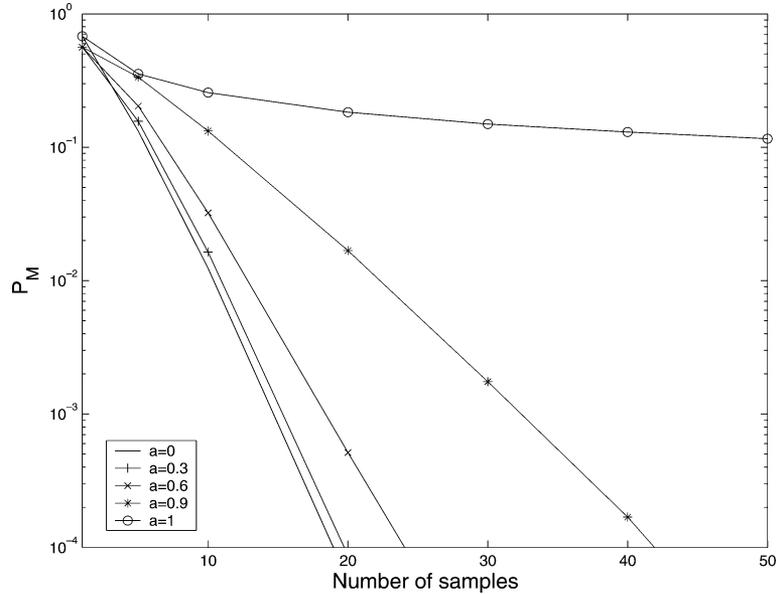


Fig. 5. P_M versus number of samples (SNR = 10 dB).

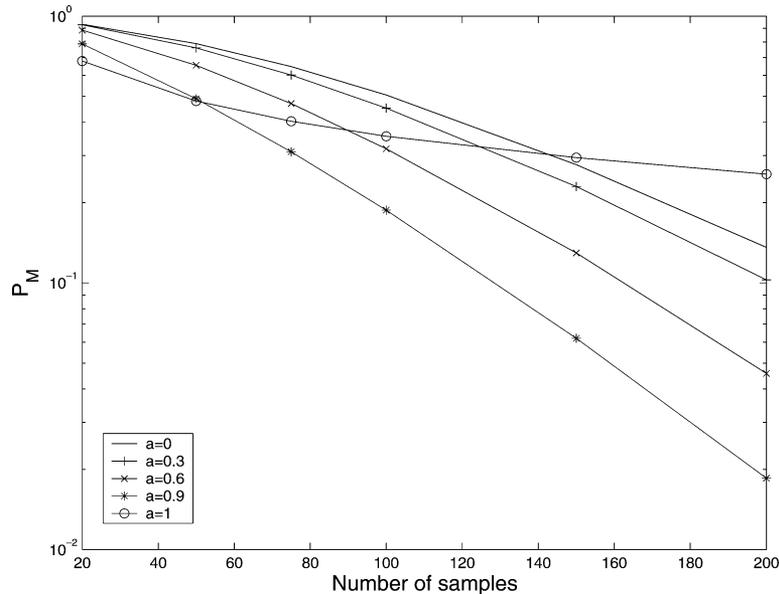


Fig. 6. P_M versus number of samples (SNR = -3 dB).

VII. CONCLUSION

We have considered the detection of correlated random signals using noisy observations. We have derived the error exponent for the Neyman–Pearson detector of a fixed level using the spectral domain and the innovations approaches. We have also provided the error exponent in closed form for the vector state-space model. The closed-form expression is valid not only for the state-space model but also for any orthogonal transformation of the original observations under the state-space model, since the spectral domain result does not change by orthogonal transformation and Theorem 1 is a closed-form expression of the invariant spectral form. We have investigated the properties of the error exponent for the scalar case. The error exponent is a function of SNR and correlation strength. The behavior of the error exponent as a function of correlation strength is sharply divided into two regimes depending on SNR. For $\text{SNR} \geq 1$, the error exponent

is monotonically decreasing in the signal correlation. On the other hand, for $\text{SNR} < 1$, there is a nonzero correlation strength that gives the maximal error exponent. Simulations confirm the validity of our asymptotic results for finite sample sizes.

APPENDIX

Proof of Theorem 1

Since the error exponent for the Neyman–Pearson detector with a fixed level $\alpha \in (0, 1)$ is given by the almost-sure limit of the normalized LLR $\frac{1}{n} \log L_n(\mathbf{y}_n)$ under H_0 (if the limit exists) [31]–[34], we focus on the calculation of the limit. We show that $\frac{1}{n} \log L_n$ converges almost surely (a.s.) under H_0 for Gauss–Markov signals in noise using the limit distribution of the innovations sequence. The LLR is given by

$$\log L_n(\mathbf{y}_n) = \log p_{1,n}(\mathbf{y}_n) - \log p_{0,n}(\mathbf{y}_n). \quad (35)$$

We have, for the second term on the right-handed side (RHS) of (35)

$$p_{0,n}(\mathbf{y}_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n y_i^2 / \sigma^2}$$

and so, under H_0

$$\frac{1}{n} \log p_{0,n}(\mathbf{y}) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2 n} \sum_{i=1}^n y_i^2 \quad (36)$$

$$\rightarrow -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \quad \text{a.s.} \quad (37)$$

since $\frac{1}{n} \sum_{i=1}^n y_i^2 \rightarrow \mathbb{E}_0\{Y_1^2\} = \sigma^2$ a.s. under H_0 as $n \rightarrow \infty$ by the SLLN. Now consider the first term on the RHS of (35). The log-likelihood under H_1 can be obtained via the Kalman recursion for the innovations [1], [6]. Specifically, define $l_i \triangleq \log p_{1,i}(y_1, y_2, \dots, y_i)$ and $\mathbf{y}_i^1 \triangleq \{y_1, y_2, \dots, y_i\}$. Then

$$p_{1,i}(y_i^1) = p_{1,i}(y_i^{i-1}) p_{1,i}(y_i | y_i^{i-1}), \quad 2 \leq i \leq n. \quad (38)$$

Hence,

$$l_i = l_{i-1} + \log p_{1,i}(y_i | y_i^{i-1}), \quad 2 \leq i \leq n \quad (39)$$

where $l_1 = \log p_{1,1}(y_1)$. Since the joint distribution of $\{y_1, y_2, \dots, y_i\}$ is Gaussian, the conditional distribution $p_{1,i}(y_i | y_i^{i-1})$ is also Gaussian with mean $\hat{y}_{i|i-1}$ and variance $R_{e,i}$. l_i is expressed using the innovations representation by

$$l_i = l_{i-1} - \frac{1}{2} \log(2\pi R_{e,i}) - \frac{1}{2} \frac{e_i^2}{R_{e,i}} \quad (40)$$

where the MMSE prediction $\hat{y}_{i|i-1}$ of y_i is the conditional expectation $\mathbb{E}_1\{y_i | y_i^{i-1}\}$ and the innovation is given by $e_i \triangleq y_i - \hat{y}_{i|i-1}$ with variance $R_{e,i} = \mathbb{E}_1\{e_i^2\}$. Hence,

$$\begin{aligned} \frac{1}{n} \log p_{1,n}(\mathbf{y}_n) &= \frac{l_n}{n} = -\frac{1}{2} \log(2\pi) \\ &\quad - \frac{1}{2n} \sum_{i=1}^n \log R_{e,i} - \frac{1}{2n} \sum_{i=1}^n \frac{e_i^2}{R_{e,i}}. \end{aligned} \quad (41)$$

The second term on the RHS of (41) is not random, and we have

$$\frac{1}{2n} \sum_{i=1}^n \log R_{e,i} \rightarrow \frac{1}{2} \log R_e \quad (42)$$

by the Cesàro mean theorem since $R_{e,i} \rightarrow R_e$ and $R_{e,i} \geq \sigma^2 > 0$ for all i where R_e is given by

$$R_e = P + \sigma^2 \quad (43)$$

and where P is the steady-state error variance of the optimal one-step predictor for the signal $\{s_i\}$. Now representing e_i as a linear combination of y_1, y_2, \dots, y_i gives

$$\begin{aligned} e_i &= y_i - K_{p,i-1} y_{i-1} - (a - K_{p,i-1}) K_{p,i-2} y_{i-2} \\ &\quad - (a - K_{p,i-1})(a - K_{p,i-2}) K_{p,i-3} y_{i-3} \\ &\quad - (a - K_{p,i-1})(a - K_{p,i-2})(a - K_{p,i-3}) K_{p,i-4} y_{i-4} - \dots \end{aligned} \quad (44)$$

where $K_{p,i} \triangleq a P_i R_{e,i}^{-1}$ is the Kalman prediction gain, $P_i \triangleq \mathbb{E}_1\{(s_i - \hat{s}_{i|i-1})^2\}$ is the error variance at time i , $\hat{s}_{i|i-1}$ is the linear MMSE prediction of s_i given y_1^{i-1} . Since the Kalman filter converges asymptotically to the time-invariant recursive Wiener filter for $0 \leq a < 1$, we have asymptotically

$$\begin{aligned} e_i &= y_i - K_p y_{i-1} - (a - K_p) K_p y_{i-2} - (a - K_p)(a - K_p) K_p y_{i-3} \\ &\quad - (a - K_p)(a - K_p)(a - K_p) K_p y_{i-4} - \dots \end{aligned} \quad (45)$$

where K_p is the steady-state Kalman prediction gain. Thus, under H_0 , the innovations sequence becomes the output of a stable recursive filter driven by an i.i.d. sequence $\{y_i\}$, and it is known to be an ergodic sequence. By the ergodic theorem, $\frac{1}{n} \sum_{i=1}^n e_i^2$ converges to the true expectation, which is given by

$$\tilde{R}_e = \lim_{i \rightarrow \infty} \mathbb{E}_0\{e_i^2\} = \sigma^2 \left(1 + \frac{K_p^2}{1 - (a - K_p)^2} \right) \quad (46)$$

since $\{y_i\}_{i=1}^{\infty}$ is an independent sequence under H_0 . Substituting K_p and R_e , we have

$$\tilde{R}_e = \sigma^2 \left(1 + \frac{a^2 P^2}{R_e^2 - a^2 \sigma^4} \right) = \sigma^2 \left(1 + \frac{a^2 P^2}{P^2 + 2\sigma^2 P + (1 - a^2)\sigma^4} \right). \quad (47)$$

Now, the last term on the RHS of (41) is given by

$$\frac{1}{2n} \sum_{i=1}^n \frac{e_i^2}{R_{e,i}} = \frac{1}{2n} \sum_{i=1}^n \frac{e_i^2}{R_e} \frac{R_e}{R_{e,i}} = \frac{1}{2n} \sum_{i=1}^n \frac{e_i^2}{R_e} \frac{R_e + C\epsilon^n - C\epsilon^n}{R_e + C\epsilon^n} \quad (48)$$

$$= \frac{1}{2n R_e} \sum_{i=1}^n e_i^2 - \frac{1}{2n R_e} \sum_{i=1}^n e_i^2 \frac{C\epsilon^n}{R_e + C\epsilon^n} \quad (49)$$

where some positive constant C and $|\epsilon| < 1$ by the exponential convergence of $R_{e,i}$ to R_e . The first term in (49) converges to $\frac{\tilde{R}_e}{2R_e}$ and the second term converges to zero since $\sum_{i=1}^n e_i^2/n$ converges to a finite constant and $R_e \geq \sigma^2 > 0$. Hence, (21) follows for $0 \leq a < 1$. When $a = 1$, we have $P_M \sim \Theta\left(\frac{1}{\sqrt{n}}\right)$ and $K = 0$. We also have $P = 0$, $\tilde{R}_e = R_e = \sigma^2$ in (22)–(24) at $a = 1$. Thus, (20) has a value of zero at $a = 1$, and Theorem 1 holds for $0 \leq a \leq 1$.

Now we show that (21) is equivalent to the spectral domain result (19) using spectral factorization. From the spectral domain form (19) we have

$$\begin{aligned} K &= \frac{1}{4\pi} \int_0^{2\pi} \log \frac{S_y^{(1)}(\omega)}{\sigma^2} d\omega + \frac{1}{4\pi} \int_0^{2\pi} \frac{\sigma^2}{S_y^{(1)}(\omega)} d\omega - \frac{1}{2}, \\ &= \frac{1}{4\pi} \int_0^{2\pi} \log S_y^{(1)}(\omega) d\omega + \frac{1}{4\pi} \int_0^{2\pi} \frac{\sigma^2}{S_y^{(1)}(\omega)} d\omega - \frac{1}{2} \log \sigma^2 - \frac{1}{2}. \end{aligned} \quad (50)$$

First, consider the first term on the RHS of (50). The argument of the logarithm is the power spectral density of the observation sequence $\{y_i\}$ under H_1 . From Wiener filtering theory, the canonical spectral factorization for $S_y^{(1)}(z)$ is given by ([2, p. 275])

$$S_y^{(1)}(z) = L(z) R_e L^*(z^{-*}) \quad (51)$$

where $L^{-1}(z)$ is the whitening filter. Hence, we have

$$\begin{aligned} & \frac{1}{4\pi} \int_0^{2\pi} \log S_y^{(1)}(\omega) d\omega \\ &= \frac{1}{4\pi} \int_0^{2\pi} \log(L(e^{j\omega})R_e L^*(e^{j\omega})) d\omega, \\ &= \frac{1}{4\pi} \int_0^{2\pi} (\log R_e + \log L(e^{j\omega}) + \log L^*(e^{j\omega})) d\omega, \\ &= \frac{1}{2} \log R_e \end{aligned}$$

where the last step follows from the cancellation of two other terms in para-Hermitian conjugacy. Now, consider the second term on the RHS of (50). From (51), we have

$$\frac{\sigma^2}{S_y^{(1)}(z)} = \frac{\sigma^2 L^{-1}(z)(L^*(z^{-*}))^{-1}}{R_e} \quad (52)$$

which is the spectral density of the innovations process under H_0 divided by R_e , since $\{y_i\}$ is an i.i.d. sequence with variance σ^2 under H_0 and $L^{-1}(z)$ is the whitening filter. Since the variance of a stationary process is given by the autocovariance function $r(l)$ setting $l = 0$, we have, by the definition of \tilde{R}_e

$$\tilde{R}_e = r(0) = \frac{1}{2\pi} \int_0^{2\pi} \sigma^2 [L^{-1}(z)(L^*(z^{-*}))^{-1}]_{z=e^{j\omega}} d\omega \quad (53)$$

since the spectral density is the Fourier transform of the autocovariance function. ((23) is an explicit formula for (53).) Hence, we have

$$\frac{1}{4\pi} \int_0^{2\pi} \frac{\sigma^2}{S_y^{(1)}(\omega)} d\omega = \frac{1}{2} \frac{\tilde{R}_e}{R_e} \quad (54)$$

and (50) is given by

$$\frac{1}{2} \log R_e + \frac{1}{2} \frac{\tilde{R}_e}{R_e} - \frac{1}{2} \log \sigma^2 - \frac{1}{2} = \frac{1}{2} \log \frac{R_e}{\sigma^2} + \frac{1}{2} \frac{\tilde{R}_e}{R_e} - \frac{1}{2}, \quad (55)$$

which is the error exponent in Theorem 1 that we derived using the innovations approach. \square

Lemma 1: The partial derivative of the error exponent with respect to the correlation coefficient a is given by

$$\frac{\partial K}{\partial a} = \frac{\Gamma(b-a)}{r_e(1-b^2)} \left(\frac{1}{1-ab} - \frac{2(1-ab)}{r_e(1-b^2)^2} \right) \quad (56)$$

for a fixed SNR Γ , where $b = a/r_e$ and $r_e = R_e/\sigma^2$.

It is easily seen that the partial derivative $\frac{\partial K}{\partial a}$ is a continuous function of a for $0 \leq a < 1$ since r_e is a continuous function of a from (22) and (24).

Proof of Lemma 1: We use the spectral domain form for the error exponent

$$\begin{aligned} K &= -\frac{1}{4\pi} \int_0^{2\pi} \log \frac{\sigma^2}{\sigma^2 + S_s(\omega)} d\omega + \frac{1}{4\pi} \int_0^{2\pi} \frac{\sigma^2}{\sigma^2 + S_s(\omega)} d\omega - \frac{1}{2} \\ &= -\frac{1}{4\pi} \int_0^{2\pi} \log \frac{1}{1 + \Gamma \tilde{S}_s(\omega)} d\omega + \frac{1}{4\pi} \int_0^{2\pi} \frac{1}{1 + \Gamma \tilde{S}_s(\omega)} d\omega - \frac{1}{2} \end{aligned} \quad (57)$$

where

$$S_s(\omega) = \frac{\Pi_0(1-a^2)}{1-2a\cos\omega+a^2}, \quad \tilde{S}_s(\omega) = S_s(\omega)/\Pi_0, \quad \Gamma = \frac{\Pi_0}{\sigma^2}. \quad (58)$$

The spectral density of the observation sequence $\{y_i\}$ is given by

$$S_y^{(1)}(z) = \sigma^2 + S_s(z) = \sigma^2 \left(1 + \frac{Q/\sigma^2}{(1-az^{-1})(1-az)} \right) \quad (59)$$

where $Q = \Pi_0(1-a^2)$, and its canonical spectral factorization is given by ([2, p. 242])

$$S_y^{(1)}(z) = \sigma^2 L(z) r_e L^*(z^{-*}) = \sigma^2 r_e \frac{1-bz}{1-az} \frac{1-bz^{-1}}{1-az^{-1}} \quad (60)$$

where $b = a/r_e$ ($|a| < 1$ and $|b| < 1$) and

$$r_e = \frac{\sqrt{[1+a^2+Q/\sigma^2]^2 - 4a^2} + 1 + a^2 + Q/\sigma^2}{2} = \frac{R_e}{\sigma^2}. \quad (61)$$

The partial derivative of K with respect to a is given by

$$\frac{\partial K}{\partial a} = \frac{1}{4\pi} \int_0^{2\pi} \frac{\Gamma \tilde{S}'_s(\omega)}{1 + \Gamma \tilde{S}_s(\omega)} d\omega - \frac{1}{4\pi} \int_0^{2\pi} \frac{\Gamma \tilde{S}'_s(\omega)}{(1 + \Gamma \tilde{S}_s(\omega))^2} d\omega \quad (62)$$

where

$$\tilde{S}'_s(\omega) = \frac{\partial \tilde{S}_s(\omega)}{\partial a} = \frac{2[(1+a^2)\cos\omega - 2a]}{(1-2a\cos\omega+a^2)^2}. \quad (63)$$

Consider the first term on the RHS of (62). Using the canonical spectral decomposition (60), we have

$$\begin{aligned} & \frac{1}{4\pi} \int_0^{2\pi} \frac{\Gamma \tilde{S}'_s(\omega)}{1 + \Gamma \tilde{S}_s(\omega)} d\omega \\ &= \frac{1}{4\pi} \int_0^{2\pi} \frac{\frac{2\Gamma[(1+a^2)\cos\omega-2a]}{(1-2a\cos\omega+a^2)^2}}{r_e \frac{1-be^{j\omega}}{1-ae^{j\omega}} \frac{1-be^{-j\omega}}{1-ae^{-j\omega}}} d\omega \\ &= \frac{1}{4\pi} \oint \frac{\Gamma[(1+a^2)(z+z^{-1})-4a]}{r_e(1-az)(1-az^{-1})(1-bz)(1-bz^{-1})} \frac{dz}{jz} \\ &= \frac{1}{4\pi j} \oint \frac{\Gamma[(1+a^2)(z^2+1)-4az]}{r_e(1-az)(z-a)(1-bz)(z-b)} dz \\ &= \frac{1}{4\pi j} \frac{\Gamma}{r_e} 2\pi j \sum_{|z|<1} \text{Residues of integrand} \\ &= \frac{\Gamma}{2r_e} \left(\frac{1-a^2}{(1-ab)(a-b)} + \frac{(1+a^2)(1+b^2)-4ab}{(1-ab)(b-a)(1-b^2)} \right) \\ &= \frac{\Gamma}{r_e} \frac{(b-a)}{(1-ab)(1-b^2)} \end{aligned} \quad (64)$$

where we have substituted $z = e^{j\omega}$, and used the residue theorem. The second term on the RHS of (62) is similarly obtained

$$\begin{aligned} & \frac{1}{4\pi} \int_0^{2\pi} \frac{\Gamma \tilde{S}'_s(\omega)}{(1 + \Gamma \tilde{S}_s(\omega))^2} d\omega = \frac{1}{4\pi} \oint \frac{\Gamma[(1+a^2)(z+z^{-1})-4a]}{r_e^2(1-bz)^2(1-bz^{-1})^2} \frac{dz}{jz} \\ &= \frac{\Gamma}{4\pi r_e^2 j} \oint \frac{(1+a^2)(z^2+1)-4az}{(1-bz)^2(z-b)^2} dz \\ &= \frac{\Gamma}{4\pi r_e^2 j} 2\pi j \text{Res}(z=b) \end{aligned}$$

where

$$\text{Res}(z=b) = \left[\frac{\partial}{\partial z} \frac{(1+a^2)(z^2+1)-4az}{(1-bz)^2} \right]_{z=b} = \frac{4(b-a)(1-ab)}{(1-b^2)^3}. \quad (65)$$

Hence, K' is given by

$$\frac{\partial K}{\partial a} = \frac{\Gamma(b-a)}{r_e(1-b^2)} \left(\frac{1}{1-ab} - \frac{2(1-ab)}{r_e(1-b^2)^2} \right). \quad (66)$$

□

Proof of Theorem 2

First, the continuity of K in (21) is straightforward as a function of R_e and \tilde{R}_e since $R_e \geq \sigma^2$, and the continuity of R_e and \tilde{R}_e is also trivial as a function of a and P from (22) and (23). Thus, we need only to show the continuity of P as a function of a , i.e., the nonnegativity of the argument of the square root in (24). The argument can be rewritten as

$$[\sigma^2(1-a^2)-Q]^2 + 4\sigma^2Q = \sigma^4(1-a^2)[(1-a^2)(1-\Gamma)^2 + 4\Gamma] \quad (67)$$

which is nonnegative if either $\Gamma = 1$ or

$$\frac{-4\Gamma}{(1-\Gamma)^2} \leq \min_{a \in [0,1]} (1-a^2) = 0$$

if $\Gamma \neq 1$. Thus, K is a continuous function of a ($0 \leq a \leq 1$) for any SNR $\Gamma \geq 0$.

i) SNR ≥ 1 :

Since $b = a/r_e$ in (66), we have

$$\begin{aligned} \frac{\partial K}{\partial a} &= \frac{\Gamma(b-a)}{r_e(1-b^2)} \left(\frac{1}{1-ab} - \frac{2b(1-ab)}{a(1-b^2)^2} \right), \\ &= \frac{\Gamma(b-a)}{r_e(1-b^2)} \frac{[a(1-b^2)^2 - 2b(1-ab)^2]}{a(1-ab)(1-b^2)^2}, \\ &= \frac{\Gamma(b-a)}{r_e(1-b^2)} \frac{[a(1+b^2)^2 - 2b(1+a^2b^2)]}{a(1-ab)(1-b^2)^2}, \\ &= \frac{\Gamma(b-a)[br_e(1+b^2)^2 - 2b(1+a^2b^2)]}{r_e a(1-ab)(1-b^2)^3}, \\ &= \frac{\Gamma(b-a)br_e^{-1}[(1+\tilde{Q}+a^2)^2 - 2r_e(1+a^4r_e^{-2})]}{r_e a(1-ab)(1-b^2)^3} \end{aligned} \quad (68)$$

where $\tilde{Q} \triangleq \Gamma(1-a^2)$ and we have used the relation $r_e(1+b^2) = 1 + \tilde{Q} + a^2$ in the canonical spectral factorization. (See [2, p. 242].) We also have the relation

$$r_e + \frac{a^2}{r_e} = 1 + \tilde{Q} + a^2 \quad (69)$$

which implies

$$r_e(1+a^4r_e^{-2}) \leq 1 + \tilde{Q} + a^2 \quad (70)$$

since $0 \leq a < 1$. Hence, for the last term in the numerator of (68) we have

$$(1+\tilde{Q}+a^2)^2 - 2r_e(1+a^4r_e^{-2}) \geq (1+\tilde{Q}+a^2)^2 - 2(1+\tilde{Q}+a^2). \quad (71)$$

The RHS of (71) is positive for $1+\tilde{Q}+a^2 = 1+a^2+\Gamma(1-a^2) \geq 2$, which reduces to the condition $\Gamma \geq 1$. Since

$$r_e = R_e/\sigma^2 = 1 + P/\sigma^2 > 1, \quad \text{for } 0 \leq a < 1 \quad (72)$$

we have

$$b-a = (r_e^{-1} - 1)a < 0. \quad (73)$$

Hence, $\frac{\partial K}{\partial a} \leq 0$ for $0 \leq a < 1$ and $\Gamma \geq 1$, and K is monotonically decreasing as $a \uparrow 1$ for $\Gamma \geq 1$.

ii) SNR < 1 :

For a given Γ , denote the last term in the numerator of (68) by

$$f_\Gamma(a) \triangleq (1 + \tilde{Q} + a^2)^2 - 2r_e(1 + a^4r_e^{-2}). \quad (74)$$

Then, we can write

$$f_\Gamma(a=0) = (1 + \Gamma)^2 - 2(1 + \Gamma) = \Gamma^2 - 1 \quad (75)$$

since $\tilde{Q} = \Gamma(1-a^2)$ and $r_e = 1 + \Gamma$ for $a = 0$ from (61). We have $f_\Gamma(a=0) < 0$ for $\Gamma < 1$ and $\frac{\partial K}{\partial a} > 0$ from (68) since $b-a < 0$. Hence, K increases as a increases in the neighborhood of $a = 0$ with $K(a=0) = D(\mathcal{N}(0,1) \parallel \mathcal{N}(0,1+\Gamma)) > 0$ if $0 < \Gamma < 1$. However, $K \downarrow 0$ as a approaches one since $P_M \sim \Theta(\frac{1}{\sqrt{n}})$ at $a = 1$. Hence, K achieves a maximum at nonzero a for SNR $\Gamma < 1$ since K is a continuous function of a , and the value of a achieving the maximum is given by $f_\Gamma(a) = 0$ since $\frac{\partial K}{\partial a}$ is also continuous with a .

As SNR $\Gamma \downarrow 0$, we have

$$\tilde{Q} = \Gamma(1-a^2) \downarrow 0 \quad \text{and} \quad r_e = 1 + P/\sigma^2 \downarrow 1. \quad (76)$$

The last term in the numerator of (68) is given by

$$(1 + \tilde{Q} + a^2)^2 - 2r_e(1 + a^4r_e^{-2}) \rightarrow (1 + a^2)^2 - 2(1 + a^4), \\ = -(1 - a^2)^2 < 0 \quad (77)$$

for $0 \leq a < 1$ as $\Gamma \downarrow 0$. Hence, for any $\delta > 0$, there exists Γ_0 small enough such that for all $\Gamma < \Gamma_0(\delta)$

$$(1 + \tilde{Q} + a^2)^2 - 2r_e(1 + a^4r_e^{-2}) \leq -(1 - a^2)^2 + \delta.$$

This guarantees that for $\Gamma < \Gamma_0(\delta)$, $\partial K/\partial a > 0$ for $0 < a < \sqrt{1 - \sqrt{\delta}}$, and $a^* \geq \sqrt{1 - \sqrt{\delta}}$. □

Proof of Theorem 3

Let $s = 1 + \Gamma \tilde{S}_s(\omega)$ where $\tilde{S}_s(\omega)$ is given by (58). Then, from (57), the partial derivative of K w.r.t. Γ is given by

$$\frac{\partial K}{\partial \Gamma} = \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial}{\partial s} \left(-\frac{1}{2} \log \frac{1}{s} + \frac{1}{2s} - \frac{1}{2} \right) \frac{\partial s}{\partial \Gamma} d\omega \quad (78)$$

where

$$\frac{\partial}{\partial s} \left(-\frac{1}{2} \log \frac{1}{s} + \frac{1}{2s} - \frac{1}{2} \right) = \frac{1}{2} \frac{s-1}{s^2} = \frac{1}{2} \frac{\Gamma \tilde{S}_s(\omega)}{s^2} > 0 \quad (79)$$

and

$$\frac{\partial s}{\partial \Gamma} = \tilde{S}_s(\omega) = \frac{1 - a^2}{1 - 2a \cos(\omega) + a^2} > 0 \quad (80)$$

$$\begin{aligned}
e_i = & y_i - \mathbf{h}^T \mathbf{K}_{p,i-1} y_{i-1} - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_{p,i-1} \mathbf{h}^T) \mathbf{K}_{p,i-2} y_{i-2} \\
& - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_{p,i-1} \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_{p,i-2} \mathbf{h}^T) \mathbf{K}_{p,i-3} y_{i-3} \\
& - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_{p,i-1} \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_{p,i-2} \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_{p,i-3} \mathbf{h}^T) \mathbf{K}_{p,i-4} y_{i-4} - \dots
\end{aligned} \quad (86)$$

for $0 \leq a < 1$. Hence,

$$\frac{\partial K}{\partial \Gamma} > 0 \quad (81)$$

and the error exponent K increases monotonically as SNR increases for a given a ($0 \leq a < 1$).

At high SNR, we have

$$\begin{aligned}
P & \sim Q \\
R_e & \sim Q + \sigma^2 \\
\tilde{R}_e & \sim \sigma^2(1 + a^2).
\end{aligned}$$

Hence, from (21), the error exponent is given at high SNR by

$$\begin{aligned}
K & \sim \frac{1}{2} \log \frac{Q + \sigma^2}{\sigma^2} + \frac{1}{2} \frac{\sigma^2(1 + a^2)}{Q + \sigma^2} \\
& = \frac{1}{2} \log \frac{Q + \sigma^2}{\sigma^2} + \frac{1}{2} \frac{1 + a^2}{\frac{Q + \sigma^2}{\sigma^2}}, \quad Q = \Pi_0(1 - a^2) \\
& = \frac{1}{2} \log \left(1 + \frac{\Pi_0}{\sigma^2} (1 - a^2) \right) + \frac{1}{2} \frac{1 + a^2}{1 + \frac{\Pi_0}{\sigma^2} (1 - a^2)}. \quad (82)
\end{aligned}$$

Since the first term is dominant at high SNR, the theorem follows. \square

Proof of Theorem 4

Since the error exponent is given by the asymptotic Kullback–Leibler rate (3) and its representation by innovations for the vector case is the same as (36) and (41). We need only to calculate R_e and \tilde{R}_e for the vector model.

The steady-state variance R_e for the innovations under H_1 is given by the conventional result of the state-space model

$$R_e = \mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma^2 \quad (83)$$

and \mathbf{P} is the unique Hermitian solution of the discrete-time Riccati equation

$$\mathbf{P} = \mathbf{A} \mathbf{P} \mathbf{A}^T + \mathbf{B} \mathbf{Q} \mathbf{B}^T - \frac{\mathbf{A} \mathbf{P} \mathbf{h} \mathbf{h}^T \mathbf{P} \mathbf{A}^T}{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma^2} \quad (84)$$

such that $\mathbf{A} - \mathbf{K}_p \mathbf{h}^T$ is stable (the existence of the solution is guaranteed since \mathbf{A} is stable, $\text{diag}(\mathbf{Q}, \sigma^2) \geq 0$, and $S_y^{(1)}(\omega) > 0$ due to the additive noise (see [2, p. 277]), where

$$\mathbf{K}_p = \mathbf{A} \mathbf{P} \mathbf{h} R_e^{-1}. \quad (85)$$

For \tilde{R}_e , we again represent e_i as a linear combination of y_1, y_2, \dots, y_i , and e_i is given by (86) at the top of the page, where $\mathbf{K}_{p,i}$ is the Kalman prediction gain given by

$$\mathbf{K}_{p,i} = \mathbf{A} \mathbf{P}_i \mathbf{h} / (\mathbf{h}^T \mathbf{P}_i \mathbf{h} + \sigma^2)$$

with the one-step prediction error covariance matrix \mathbf{P}_i [2]. Since the Kalman filter converges asymptotically to the time-invariant recursive Wiener filter when \mathbf{A} is stable, we have asymptotically

$$\begin{aligned}
e_i = & y_i - \mathbf{h}^T \mathbf{K}_p y_{i-1} - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) \mathbf{K}_p y_{i-2} \\
& - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) \mathbf{K}_p y_{i-3} \\
& - \mathbf{h}^T (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) \mathbf{K}_p y_{i-4} - \dots
\end{aligned} \quad (87)$$

where \mathbf{K}_p is the steady-state Kalman prediction gain. Thus, the innovation sequence becomes the output of a stable recursive filter driven by the i.i.d. sequence $\{y_i\}$ under H_0 as in the scalar case, and the ergodic theorem holds for $\frac{1}{n} \sum_{i=1}^n e_i^2$.

$$\begin{aligned}
\tilde{R}_e & = \lim_{i \rightarrow \infty} \mathbb{E}_0 \{e_i^2\}, \\
& = \sigma^2 + \sigma^2 \mathbf{h}^T \left(\sum_{k=0}^{\infty} (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^k \mathbf{K}_p \mathbf{K}_p^T [(\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^k]^T \right) \mathbf{h}.
\end{aligned} \quad (88)$$

Let $\tilde{\mathbf{P}}$ be defined as

$$\tilde{\mathbf{P}} \triangleq \sum_{k=0}^{\infty} (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^k \mathbf{K}_p \mathbf{K}_p^T [(\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^k]^T. \quad (89)$$

$\tilde{\mathbf{P}}$ is finite since $\mathbf{A} - \mathbf{K}_p \mathbf{h}^T$ is stable by the property of the solution of (84), and is given by the unique solution of the following Lyapunov equation:

$$\tilde{\mathbf{P}} - (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T) \tilde{\mathbf{P}} (\mathbf{A} - \mathbf{K}_p \mathbf{h}^T)^T = \mathbf{K}_p \mathbf{K}_p^T. \quad (90)$$

(Since $\mathbf{A} - \mathbf{K}_p \mathbf{h}^T$ is stable and $\mathbf{K}_p \mathbf{K}_p^T \geq 0$, there exists a unique, Hermitian, and positive semi-definite solution $\tilde{\mathbf{P}}$ of (90) [2].) The spectrum for the vector model is given by (34) [2]. \square

ACKNOWLEDGMENT

The authors wish to thank an anonymous reviewer for pointing out a simple derivation of the spectral domain result in Section III.

REFERENCES

- [1] C. W. Helstrom, *Elements of Signal Detection and Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [2] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
- [3] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [4] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.
- [5] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2230–2259, Oct. 1998.

- [6] F. C. Schwappe, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 61–70, Jan. 1965.
- [7] H. L. van Trees, "Applications of state-variable techniques in detection theory," *Proc. IEEE*, vol. 58, no. 5, pp. 653–669, May 1970.
- [8] F. C. Schwappe, "State space evaluation of the Bhattacharyya distance between two Gaussian processes," *Inf. Contr.*, vol. 11, no. 3, pp. 352–372, 1967.
- [9] —, "On the Bhattacharyya distance and the divergence between Gaussian processes," *Inf. Contr.*, vol. 11, no. 4, pp. 373–395, 1967.
- [10] T. Kailath, "The divergence and Bhattacharyya distance measures in signal detection," *IEEE Trans. Commun. Technol.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.
- [11] —, "The innovations approach to detection and estimation theory," *Proc. IEEE*, vol. 58, no. 5, pp. 680–695, May 1970.
- [12] U. Grenander and G. Szegő, *Toeplitz Forms and their Applications*. Berkeley, CA: Univ. California Press, 1958.
- [13] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.
- [14] E. J. Hannan, "The asymptotic theory of linear time-series models," *J. Appl. Probab.*, vol. 10, pp. 130–145, 1973.
- [15] F. den Hollander, *Large Deviations*. Providence, RI: Amer. Math. Soc., 2000, vol. 14, Fields Institute Monographs.
- [16] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones and Bartlett, 1993.
- [17] J. Gärtner, "On large deviations from an invariant measure," *Teor. Veroyatnost. i Primenen.*, vol. 22, no. 1, pp. 27–42, 1977.
- [18] R. S. Ellis, "Large deviations for a general class of random vectors," *Ann. Probab.*, vol. 12, no. 1, pp. 1–12, Feb. 1984.
- [19] M. D. Donsker and S. R. S. Varadhan, "Large deviations for stationary Gaussian process," *Commun. Math. Phys.*, vol. 97, pp. 187–210, 1985.
- [20] G. R. Benitz and J. A. Bucklew, "Large deviation rate calculations for nonlinear detectors in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 358–371, Mar. 1990.
- [21] R. K. Bahr, "Asymptotic analysis of error probabilities for the non-zero-mean Gaussian hypothesis testing problem," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 597–607, Mar. 1990.
- [22] R. K. Bahr and J. A. Bucklew, "Optimal sampling schemes for the Gaussian hypothesis testing problem," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1677–1686, Oct. 1990.
- [23] P. Barone, A. Gigli, and M. Piccioni, "Optimal importance sampling for some quadratic forms of ARMA processes," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1834–1844, Nov. 1995.
- [24] W. Bryc and W. Smolenski, "On the large deviation principle for a quadratic functional of the autoregressive process," *Statist. and Probab. Lett.*, vol. 17, pp. 281–285, 1993.
- [25] W. Bryc and A. Dembo, "Large deviations for quadratic functionals of Gaussian processes," *J. Theor. Probab.*, vol. 10, no. 2, pp. 307–332, 1997.
- [26] B. Bercu, F. Gamboa, and A. Rouault, "Large deviations for quadratic forms of stationary Gaussian processes," *Stochastic Processes and their Applications*, vol. 71, pp. 75–90, 1997.
- [27] B. Bercu and A. Rouault, "Sharp large deviations for the Ornstein-Uhlenbeck process," *Theory of Probability and its Applications*, vol. 46, no. 1, pp. 1–19, 2002.
- [28] J.-F. Chamberland and V. V. Veeravalli, "Design of sensor networks for detection applications via large-deviation theory," in *Proc. IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004, pp. 153–158.
- [29] R. R. Bahadur, *Some Limit Theorems in Statistics*. Philadelphia, PA: SIAM, 1971, vol. 4, Conference Board of the Mathematical Sciences Regional Conferences Series in Applied Mathematics.
- [30] R. R. Bahadur, S. L. Zabell, and J. C. Gupta, "Large deviations, tests, and estimates," in *Asymptotic Theory of Statistical Tests and Estimation: In Honor of Wassily Hoeffding*, I. M. Chakravarti, Ed. New York: Academic, 1980.
- [31] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, The Netherlands: Kluwer, 1989.
- [32] —, "Distances and discrimination rates of stochastic processes," *Stochastic Processes and their Applications*, vol. 35, pp. 47–57, 1990.
- [33] P.-N. Chen, "General formulas for the Neyman-Pearson type-II error exponent subject to fixed and exponential type-I error bounds," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 316–323, Jan. 1996.
- [34] H. Luschgy, A. L. Rukhin, and I. Vajda, "Adaptive tests for stochastic processes in the ergodic case," *Stochastic Processes and their Applications*, vol. 45, pp. 47–59, 1990.
- [35] H. Luschgy, "Asymptotic behavior of Neyman-Pearson tests for autoregressive processes," *Scand. J. Statist.*, vol. 21, pp. 461–473, 1994.