where x_0 is the 9900th largest value of the statistic. Thresholds were estimated for false alarm probabilities $P_F = 10^{-k}$, $k = 2, 3, 4, 5, 6$ for each repetition of the experiment. Histograms of these threshold values are shown in Figs. 3 and 4, for $P_F = 10^{-2}$ and $P_F = 10^{-4}$. Since the underlying distribution of $L(\cdot)$ is unknown, one measure of the accuracy of the estimate is the extent to which most of the estimates fall in one bin of the histogram. Also, we can see that there is no overlap between the estimated threshold values in the histograms for the two different P_F's. This supports the claim that the estimated threshold is likely to yield a false alarm probability which is of the same order as the desired P_F. A small overlap was noticed in the thresholds of the histograms for $P_F = 10^{-5}$ and $P_F = 10^{-6}$. Also, there is much higher spread in the threshold values estimated for lower false alarm probabilities. Based on the excellent results obtained for the same choices of P_F's in the known cases of the previous section, these results are surprising. However, it is explained as follows. The γ values of the GPD estimated for the different repetitions of this experiment lie in the range 0.45–0.55. This represents an extremely heavy-tailed distribution. The lognormal distribution, which is quite a heavy-tailed distribution, was found to have $\gamma = 0.232$. The heavy-tailed nature of the detector statistic can also be observed by comparing the large threshold values seen in the histograms with the corresponding thresholds of the Gaussian and the lognormal distributions. The variance of the GPD is ∞ for $\gamma \geq 0.5$. Thus the bivariate Laplace results in a highly fluctuating statistic with an extremely large variance. As such, it represents a "worst case" situation for empirically determining the threshold. By counting the number of estimates that fell into the bins between 10 and 16 for $P_F = 10^{-4}$ it was found that 82% of the estimates fell into these bins. Thus even for this extremely large-tailed example, we believe that use of the GPD has allowed us to estimate useful values for the thresholds with sample sizes much smaller than $10/P_F$.

IV. Conclusion

In this correspondence the ordered sample least squares estimator is proposed for estimation of the generalized Pareto distribution which is used to approximate the extreme tails of a probability density function. Application of the technique to known distributions reveals that excellent results can be had with sample sizes that are orders of magnitude smaller than those required for conventional Monte Carlo techniques. In particular, the thresholds required for very low false alarm probabilities were obtained with a reasonable accuracy for both known and unknown distribution cases as is shown in the tables and examples.

References

[1] L. Breiman, C. J. Stone, and J. D. Gins, "New methods for estimating tail probabilities and extreme value distributions," Tech. Rep. TSC-PD 266-1, Technology Service Corp., Santa Monica, CA 90405.

[2] E. Conte, M. Lops, and S. Ullo, "Detection of signals with unknown parameters in correlated K-distributed clutter," *Proc. Inst. Elec. Eng.*, pt. F, vol. 138, no. 2, Apr. 1991.

[3] ——, "A new model for coherent Weibull clutter," in *Proc. Int. Conf. on Radar* (Paris, France, Apr. 24–28, 1989), pp. 482–487.

[4] A. Davison, "Modeling excesses over high threshold, with an application," in *Statistical Extremes and Applications*, Tiago de Oliviera, Ed., NATO ASI Series C, vol. 131, 1984, pp. 461–482.

[5] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique," *Annals Statist.*, vol. 11, pp. 1019–1031, 1983.

[6] W. H. DuMouchel and R. A. Olshen, "On the distributions of claim costs," in *Credibility Theory and Applications*, P. M. Kahn, Ed. New York: Academic Press, 1975, pp. 23–46.

[7] A. Farina and A. Russo, "Radar detection of correlated targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 5, pp. 513–536, Sept. 1982.

[8] J. Galambos, "Asymptotics; stable laws for extremes; tail properties," in *Statistical Extremes and Applications*, Tiago de Oliviera, Ed., NATO ASI Series C, vol. 131, 1984, pp. 19–29.

[9] M. Guida, D. Iovino, and M. Longo, "Comparitive performance analysis of some extrapolative estimators of probability tails," *IEEE J. Sel. Areas Commun.*, vol. 6, pp. 76–84, 1988.

[10] F. E. Harrel and C. E. Davis, "A new distribution-free quantile estimator," *Biometrika*, vol. 69, pp. 635–640, 1982.

[11] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation of the generalized Pareto distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

[12] S. S. Kassam, *Signal Detection in Non-Gaussian Noise*, J. Thomas, Ed. New York: Springer-Verlag, 1987.

[13] A. B. Martinez, P. F. Swaszek, and J. B. Thomas, "Locally optimal detection in multivariate non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 6, pp. 815–822, Nov. 1984.

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.

[15] I. Weissman, "Estimation of parameters and large quantiles based on the k largest observations," *J. Amer. Statist. Assoc.*, vol. 73, pp. 812–815, 1978.

Convexity Properties in Binary Detection Problems

Murat Azizoğlu, *Member, IEEE*

Abstract—**This correspondence investigates convexity properties of error probability in the detection of binary-valued scalar signals corrupted by additive noise. It is shown that the error probability of the maximum-likelihood receiver is a convex function of the signal power when the noise has a unimodal distribution. Based on this property, results on the optimal time-sharing strategies of transmitters and jammers, and on the optimal use of multiple channels are obtained.**

Index Terms— **Binary scalar signals, maximum-likelihood detection, additive-noise channels, error probability, convexity, optimum time sharing, optimum jamming strategy, channel switching.**

I. Introduction

In this correspondence we investigate convexity properties of error probability in optimum detection of digital signals. We consider binary-valued scalar signals transmitted over an arbitrary additive noise channel, and show that the error probability of the maximum-likelihood receiver is a convex function of the signal power for a broad class of noise distributions. We then explore the implications of convexity for the optimal time-sharing strategies for average-power-constrained transmitters and jammers, and for optimal switching between multiple additive noise channels.

Manuscript received August 28, 1995; revised January 29, 1996. This work was supported by the NSF under Grants NCR-9309574 and NCR-9502610. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 1994.

The author is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA.

Publisher Item Identifier S 0018-9448(96)04475-6.

Consider a scalar communication system where binary messages are encoded into signal values which are in turn corrupted by additive noise. The received scalar signal is $Y = X + N$, where X assumes values from the signal set $\{\alpha_0, \alpha_1\}$, and the additive noise N is statistically independent of the signal X. The noise is average-power constrained so that $E(N^2) \le \sigma^2$, and has an otherwise arbitrary probability density function $p_N(\cdot)$. We will assume equiprobable symbols. In Section II, we will briefly address the impact of nonuniform input distribution on convexity.

The scalar channel model above may serve as an abstraction for a continuous-time system in which the received signal is processed by a linear filter and subsequently sampled once per symbol interval. Thus the scalar model combines the modulator, the additive-noise channel, and the described front-end receiver. The receiver front-end is followed by a maximum-likelihood (ML) scalar receiver. The scalar noise N may capture the effects of noise, jamming, intersymbol interference, multiuser interference, etc. It must be noted that the continuous-time receiver will be suboptimal for non-Gaussian noise, as noise projections other than that captured by the filter are not utilized [1].

This scalar system model is used by Shamai and Verdú in [2] to obtain the least favorable noise distributions for error probability and channel capacity when the signal values are fixed at ± 1. In both cases, the worst case distribution is found to be a mixture of discrete lattices. For any discrete noise distribution, the error probability can be taken to zero via *time-sharing*; the transmitter time-shares between appropriate signal values such that there is no ambiguity at the receiver about the transmitted symbol. Since we allow such time-sharing, we will consider only continuous noise distributions in this correspondence.

The first issue we address is whether the error probability of the ML receiver is a convex function of the average signal power $S = (\alpha_0^2 + \alpha_1^2)/2$. This is motivated by the possibility of improving the error performance through time-sharing between two sets of signal power levels while satisfying an average power constraint. In Section II we show that the error probability $e(S)$ is convex when the noise has a unimodal and differentiable density function. This holds not only for the ML receiver, but also for a large class of suboptimal receivers.

A similar question relates to the concavity of the error probability with respect to the average noise power σ^2. This is motivated by the viewpoint of a jammer who is interested in the possibility of increasing the error probability through time-sharing. In Section III, we show that a critical value for the noise power determines the jammer's optimum strategy. Time-sharing aids the jammer if and only if its average power is below the critical value.

In Section IV we consider the problem of optimal channel switching. Here two additive noise channels are available, and either one can be used for a given symbol transmission. We obtain the optimum use of these channels under certain structural conditions. When the average transmitter power is between two critical values, the optimal solution uses both channels; otherwise, the better channel is used exclusively.

II. Convexity in Signal Power

The received signal under hypothesis i $(i = 0, 1)$ is $Y = \alpha_i + N$. Let P_e^i be the error probability of the ML receiver under hypothesis i. Then

$$P_e^i = \int_{A_i} p_Y(y \mid i)\, dy$$

where $p_Y(y \mid i)$ is the conditional probability density for the received signal under hypothesis i, and A_i is the set of received points which result in a symbol error. In the case of an additive-noise channel with equiprobable symbols, the error probability can be written as

$$P_e = \frac{1}{2} \int \min\left[p_N(y), p_N(y - (\alpha_1 - \alpha_0))\right] dy. \tag{1}$$

Two well-known observations are immediate from (1). First, the ML error probability is not affected by the mean value of the noise, as the mean merely shifts the integrand along the real line. Second, the error probability depends on the signal levels α_0 and α_1 only through the Euclidean distance $|\alpha_1 - \alpha_0|$. In the case where the error probability is a nonincreasing function of this distance,[1] the optimum placement of signals subject to an average power constraint $(\alpha_0^2 + \alpha_1^2)/2 \le S$ results in antipodal signaling, i.e., $\alpha_1 = \sqrt{S}$, $\alpha_0 = -\sqrt{S}$.

While the average power constraint concerns the second moment of the signal, the error probability is a function of its variance $(\alpha_1 - \alpha_0)^2/4$. For a fixed mean signal value, the error probability of a binary signaling scheme is a convex function of its power if and only if the same convexity holds for antipodal signaling. Thus we will consider antipodal signaling without loss of generality, and we denote by $e(S)$ the error probability achieved by the ML receiver with power S.

The following definition will be used in the rest of this correspondence.

Definition: The class of unimodal differentiable density functions $\mathcal{D}_u$ consists of density functions which have a single local maximum (at some point x_0) and are differentiable at every point except possibly at x_0. Formally

$$\mathcal{D}_u(x_0) = \left\{ p(x) \in \mathcal{C}'(x_0) : p(x) \ge 0, \int p(x)\, dx = 1, \int x^2 p(x)\, dx < \infty, (x - x_0)p'(x) \le 0 \quad \forall x \ne x_0 \right\}$$

and

$$\mathcal{D}_u = \bigcup_{x_0 \in \Re} \mathcal{D}_u(x_0)$$

where $\mathcal{C}'(x_0)$ is the class of continuously differentiable functions on $\Re - \{x_0\}$.

Two subclasses of $\mathcal{D}_u$ will also be of interest. The class of symmetric unimodal densities $\mathcal{D}_s$ include $p \in \mathcal{D}_u$ that satisfy $p(x) = p(2x_0 - x)$ for some $x_0 \in \Re$. The class of one-sided unimodal densities $\mathcal{D}_\pm$ include $p \in \mathcal{D}_u$ that are defined either on (x_0, ∞) or on $(-\infty, x_0)$ for some $x_0 \in \Re$, and are monotonic on the corresponding interval. The subclasses $\mathcal{D}_+$ and $\mathcal{D}_-$ refer to the first and the second type of interval, respectively. Finally, $\mathcal{D}_s(x_0)$, $\mathcal{D}_+(x_0)$, and $\mathcal{D}_-(x_0)$ are defined similarly.

Many common density functions belong to the classes just defined. The Gaussian density

$$p(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-m)^2/2\sigma^2}$$

is in $\mathcal{D}_s(m)$, as is a symmetric exponential (Laplace) density $p(x) = ae^{-2a|x-m|}$. The (one-sided) exponential density is in $\mathcal{D}_+(0)$.

Since the error probability is invariant under shifts of the noise density, we consider the class $\mathcal{D}_u(0)$ without loss of generality. Let

$$p_N(x) = \begin{cases} f(x), & x \ge 0 \\ g(-x), & x < 0 \end{cases} \tag{2}$$

where f and g are nonnegative, nonincreasing functions which are differentiable on $(0, \infty)$. The ML receiver for antipodal signaling with the noise in this class is a threshold detector; in general, the

[1] Later in this section we provide a sufficient condition for this to hold.

threshold lies in $[-\sqrt{S}, \sqrt{S}]$ and its value depends on the tails of the density function. The optimal threshold t_0 is the root of the function

$$h(x) = f(x + \sqrt{S}) - g(-x + \sqrt{S})$$

defined over $[-\sqrt{S}, \sqrt{S}]$, whenever a root exists. Note that since $h(x)$ is nonincreasing over $(-\sqrt{S}, \sqrt{S})$, multiple isolated roots cannot exist. If $h(x) > 0$ over $(-\sqrt{S}, \sqrt{S})$, then $t_0 = \sqrt{S}$. Conversely if $h(x)$ is uniformly negative, then $t_0 = -\sqrt{S}$.

The error probability is given by

$$e(S) = \frac{1}{2}\int_{t_0+\sqrt{S}}^{\infty} f(x)dx + \frac{1}{2}\int_{\sqrt{S}-t_0}^{\infty} g(x)\,dx \tag{3}$$

which upon differentiation yields

$$\begin{aligned}\frac{d}{dS}e(S) = &-\frac{1}{2}f(t_0 + \sqrt{S})\frac{d}{dS}(t_0 + \sqrt{S})\\ &- \frac{1}{2}g(\sqrt{S} - t_0)\frac{d}{dS}(\sqrt{S} - t_0).\end{aligned} \tag{4}$$

The fact that $e(S)$ is a nonincreasing function of S can be seen by considering the three possible cases. When $t_0 \in (-\sqrt{S}, \sqrt{S})$, we have

$$f(t_0 + \sqrt{S}) = g(\sqrt{S} - t_0)$$

and

$$de(S)/dS = -f(t_0 + \sqrt{S})/2\sqrt{S} \leq 0.$$

On the other hand, when $t_0 = \pm\sqrt{S}$, one of the two terms in (4) vanishes and the other term is nonpositive.

The general convexity result can now be established.

Theorem 1: $e(S)$ is a convex nonincreasing function for any noise density in $\mathcal{D}_u$.

Proof: It suffices to consider a density function in $\mathcal{D}_u(0)$ as given in (2). Then

$$\begin{aligned}\frac{d^2}{dS^2}e(S) = &-\frac{1}{2}f'(t_0 + \sqrt{S})\left[\frac{d}{dS}(t_0 + \sqrt{S})\right]^2\\ &- \frac{1}{2}g'(\sqrt{S} - t_0)\left[\frac{d}{dS}(\sqrt{S} - t_0)\right]^2\\ &- \frac{1}{2}f(t_0 + \sqrt{S})\frac{d^2}{dS^2}(t_0 + \sqrt{S})\\ &- \frac{1}{2}g(\sqrt{S} - t_0)\frac{d^2}{dS^2}(\sqrt{S} - t_0).\end{aligned} \tag{5}$$

The first two terms on the right-hand side of (5) are nonnegative. When $|t_0| < \sqrt{S}$, using $f(t_0 + \sqrt{S}) = g(\sqrt{S} - t_0)$, the last two terms can be combined as $f(t_0 + \sqrt{S})/4S^{3/2}$, which is also nonnegative. When $|t_0| = \sqrt{S}$, one of the last two terms vanish and the other is nonnegative. Therefore, $e(S)$ is convex since it has a nonnegative second derivative. □

The following corollary shows that convexity also extends to a class of suboptimal threshold detectors.

Corollary 1: The error probability of a threshold detector is convex for any noise distribution in $\mathcal{D}_u$ if its threshold $t(S)$ satisfies $|d^2t/dS^2| \leq S^{-3/2}/4$.

Proof: The error probability of the threshold detector is given by (3) with t_0 replaced by the suboptimal threshold t. From (5) it is seen that sufficient conditions for convexity are $d^2(t + \sqrt{S})/dS^2 \leq 0$ and $d^2(\sqrt{S} - t)/dS^2 \leq 0$, which result in the claim. □

The convexity in Corollary 1 is robust in that it holds for any noise distribution in $\mathcal{D}_u$ while the optimum threshold for a particular density in $\mathcal{D}_u$ may yield a nonconvex error probability for another density in $\mathcal{D}_u$. Also, a suboptimal threshold satisfying this condition results in convex conditional error probabilities P_e^0 and P_e^1, while this may not hold with the optimum threshold. Thus the ML receiver may not yield a convex error probability for nonequiprobable symbols (unlike the maximum a posterori (MAP) receiver). The convexity of the MAP error probability can be seen by replicating the proof of Theorem 1 with the modification

$$h(x) = q_0 f(x + \sqrt{S}) - q_1 g(-x + \sqrt{S})$$

where q_0 and q_1 are the *a priori* symbol probabilities. An example of nonconvexity with the ML receiver occurs with $f(x) = 1/8e^{-x/4}$, $g(x) = 1/4e^{-x/2}$, and $q_0 > 1/2$.

The convexity in Theorem 1 implies that an average power-limited transmitter cannot improve its error performance via time-sharing between different power levels. Such time-sharing would achieve a performance

$$\sum_i p_i e(S_i) \geq e\left(\sum_i p_i S_i\right) = e(S)$$

where $0 \leq p_i \leq 1$ is the fraction of the time power S_i is used and S is the average power.

Convexity can also be used to obtain a universal lower bound to the error probability in problems where the received power is random, e.g., due to amplitude fading, through Jensen's inequality [3].

Simple examples can be constructed to show that unimodality of the noise density is not a necessary condition for convexity.

For a symmetric noise density the error probability corresponds to that of a binary-symmetric channel (BSC) with crossover probability $e(S)$, while for a one-sided noise density, the corresponding model is a Z-channel with crossover probability $2e(S)$. Corollary 2, which follows from (3), compares the performance of these channels.

Corollary 2: Let $p_N(x) \in \mathcal{D}_s(0)$ have error probability $e_{\mathrm{BSC}}(S)$, and let $p_{N'}(x) = 2p_N(x)u(x) \in \mathcal{D}_+(0)$ have error probability $e_Z(S)$. Then $e_Z(S) = e_{\mathrm{BSC}}(4S)$, i.e., the BSC with noise N has a 6-dB performance disadvantage over the Z-channel with the same power S and with noise $N' = |N|$.

III. Convexity in Noise Power

In the previous section, we considered the noise power σ^2 to be fixed and investigated the possibility of improving the error performance through time-sharing of signal power. In this section we consider the dual problem of time-sharing of noise power. Here the signal power will be fixed, and the perspective is that of a jammer who is interested in maximizing the error probability subject to an average power constraint. The noise density will belong to the symmetric, zero-mean class $\mathcal{D}_s(0)$ so that the sign detector is the optimum receiver regardless of the noise power used. The jammer is not allowed to switch arbitrarily among different densities in $\mathcal{D}_s(0)$, but only to change the linear scaling of a random variable. That is, for a given symbol transmission the jammer uses power β via the noise variable $N = \sqrt{\beta}N_0$ where $E(N_0^2) = 1$ and $p_{N_0}(x) \in \mathcal{D}_s(0)$. The jammer can time-share subject to the constraint

$$\sum_i p_i\beta_i \leq \sigma^2$$

where p_i is the fraction of the time power β_i is used.

With antipodal signaling and constant transmitter power S, the received signal can be normalized as $Y = \pm 1 + N$, and the average power constraint for the jammer becomes

$$E(N^2) \leq \gamma \triangleq \frac{\sigma^2}{S}. \tag{6}$$

The error probability as a function of normalized jammer power γ is $j(\gamma) \triangleq e(\gamma^{-1})$ where $e(\cdot)$ is as in Section II. The following theorem shows that $j(\gamma)$ is neither convex nor concave.

Theorem 2: For any noise density in $\mathcal{D}_s(0)$, $j(\gamma)$ is a nondecreasing function of γ with at least one inflection point.

Proof: The noise density is of the form $p_N(x) = f(|x|)$ where f is nonincreasing and differentiable on $(0, \infty)$. Since $e(\cdot)$ is nonincreasing, $j(\cdot)$ is nondecreasing. Differentiating

$$j(\gamma) = \int_{\gamma^{-1/2}}^{\infty} f(x)\, dx$$

one obtains

$$\frac{d^2}{d\gamma^2} j(\gamma) = -\frac{1}{4}\gamma^{-5/2}[3f(\gamma^{-1/2}) + \gamma^{-1/2} f'(\gamma^{-1/2})]. \quad (7)$$

Let $y = \gamma^{-1/2}$ and let $h(y) = 3f(y) + yf'(y)$. For an arbitrary $a > 0$ one has

$$\int_0^a y^2 h(y) dy = a^3 f(a).$$

Since the noise has a finite variance, $\lim_{a\to\infty} a^3 f(a) = 0$. Then the function $y^2 h(y)$ integrates to 0 over $(0, \infty)$. Hence $h(y)$ must change sign, and, by continuity, must have at least one positive root. It follows from (7) that $j(\gamma)$ has at least one inflection point. □

One can also show, with an argument similar to that used in the proof above, that the number of inflection points of $j(\gamma)$ is odd. For a complete characterization of optimal jamming strategy, as described by Theorem 3 below, we assume that the noise density is such that $j(\gamma)$ has a single inflection point. This holds true for a wide class of symmetric unimodal densities including Gaussian and Laplace densities.[2] Then $j(\gamma)$ is convex over $(0, \gamma^*)$ and concave over (γ^*, ∞) where γ^* is the unique inflection point.

The optimal time-sharing strategy of the jammer achieves

$$j^*(\gamma) = \sup_{n, E_n} \sum_{i=1}^{n} p_i j(\gamma_i) \quad (8)$$

where

$$E_n = \left\{(p_i, \gamma_i) : i = 1, 2, \cdots, n, 0 \le p_i \le 1, \gamma_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \gamma_i = \gamma\right\}.$$

The following lemma shows that $j^*(\gamma)$ has a well-defined structure [4].

Lemma 1: $j^*(\gamma)$ is the smallest concave function that is larger than $j(\gamma)$.

From Caratheodory's theorem [4], at most two power levels are needed for optimal time-sharing. These observations are valid irrespective of the number of inflection points of $j(\gamma)$, and they can be used to construct the optimal time-sharing strategy when the assumption regarding the uniqueness of the inflection point does not hold.

To obtain the relationship between the optimal jamming strategy and the jammer power, we need an auxiliary result, proved in the Appendix, about the existence of a tangent to $j(\gamma)$ passing through the origin.

Lemma 2: Let γ^* be the inflection point of $j(\gamma)$. There exists a unique point $\gamma_c \ge \gamma^*$ such that the tangent to $j(\gamma)$ at γ_c lies above $j(\gamma)$ and passes through the origin.

For a given $j(\cdot)$, γ_c can be found as the unique solution to the equation $j(x) = xj'(x)$, for $x \ge \gamma^*$.

[2] It is of interest to note that the Cauchy density, which is not in $\mathcal{D}_s(0)$, results in a concave $j(\gamma)$.

Theorem 3: The jammer's optimal strategy, when its average power γ is below γ_c, is to switch between powers 0 and γ_c with the fraction of on-power time γ/γ_c. For $\gamma > \gamma_c$, the optimum strategy does not employ time-sharing.

Proof: The stated strategy achieves an error probability of

$$j_0(\gamma) = \begin{cases} j(\gamma_c)\gamma/\gamma_c = j'(\gamma_c)\gamma, & \gamma \le \gamma_c \\ j(\gamma), & \gamma > \gamma_c. \end{cases}$$

Since $j_0(\cdot)$ has a nonpositive second derivative, it is concave. According to Lemma 1 it suffices to show that $j_0(\gamma)$ is the smallest concave function larger than $j(\gamma)$. This is clearly the case for $\gamma > \gamma_c$. Suppose there is another concave function $f(\gamma)$ which is larger than $j(\gamma)$ and which satisfies $j_0(\gamma_0) > f(\gamma_0) \ge j(\gamma_0)$ for some $\gamma_0 \le \gamma_c$. Then for any γ_1, γ_2, and $0 \le \lambda \le 1$ such that $\lambda\gamma_1 + (1-\lambda)\gamma_2 = \gamma_0$ we have

$$f(\gamma_0) \ge \lambda f(\gamma_1) + (1-\lambda) f(\gamma_2) \ge \lambda j(\gamma_1) + (1-\lambda) j(\gamma_2).$$

Now let $\gamma_1 = \gamma_c$, $\gamma_2 = 0$ and $\lambda = \gamma_0/\gamma_c$. Then $f(\gamma_0) \ge j(\gamma_c)\gamma_0/\gamma_c = j_0(\gamma_0)$ which is a contradiction. □

Theorem 3 states that the optimal strategy of the jammer is uniquely determined by γ_c. A weak jammer should employ an on–off time-sharing with γ_c as its on-power, whereas a strong jammer should continuously operate at its average power.

A jammer employing an on–off strategy has a noise density

$$p_N(x) = (1-p)\delta(x) + \sqrt{p^3/\sigma^2} p_0(\sqrt{p}x/\sigma)$$

where $p_0(\cdot)$ is the density of the unit power noise and p is the fraction of the on-time. Interestingly, the optimal strategy of the *transmitter* does not change with jammer's time-sharing. As long as the transmitter does not know which symbols are affected by the jammer, its error probability is $pe(pS)$ which cannot be improved by time-sharing of transmitter power. In game-theoretic terminology, the constant power transmitter and on–off time-sharing jammer is an equilibrium pair [5].

The on–off jamming strategy above has received considerable attention in the spread spectrum literature where the noise has Gaussian statistics [6]–[8]. The well-known *pulsed jammer* uses an on–off strategy to maximize the error probability. The special cases of the result in Theorem 3 with spread-spectrum signals and Gaussian noise can be found in [9] and [10], and use of error correction coding to improve the performance is considered in [8] and [11]. When the channel capacity is the performance measure instead of the bit-error probability, analogous results have been obtained for the Gaussian jammer [12]. Hegde *et al.* [13] consider the time-sharing of the jammer power when the channel output is quantized. For recent results on the optimal jamming strategies with concatenated coding and parallel decoders the reader is referred to [14].

IV. Channel Switching

In this section we apply the convexity result of Section II to a system in which the transmitter/receiver pair is connected via multiple additive noise channels as depicted in Fig. 1. The transmitter has a switch which controls the access to these channels. For a given symbol transmission any one of the channels can be used. It is assumed that the receiver knows which channel is currently in use, e.g., the switches at the transmitter and the receiver are synchronized. The noise N_i on channel i has a density $p_i(x)$ which is in the class $\mathcal{D}_u$. We assume that the average noise power in each of the channels is the same so that the received signal-to-noise ratio does not depend on the channel in use. The receiver is the optimal one for the corresponding channel.

We are interested in the channel switching strategy that minimizes the average error probability $e(S)$. Let $e_i(S)$ $i = 1, 2, \cdots, M$, be the

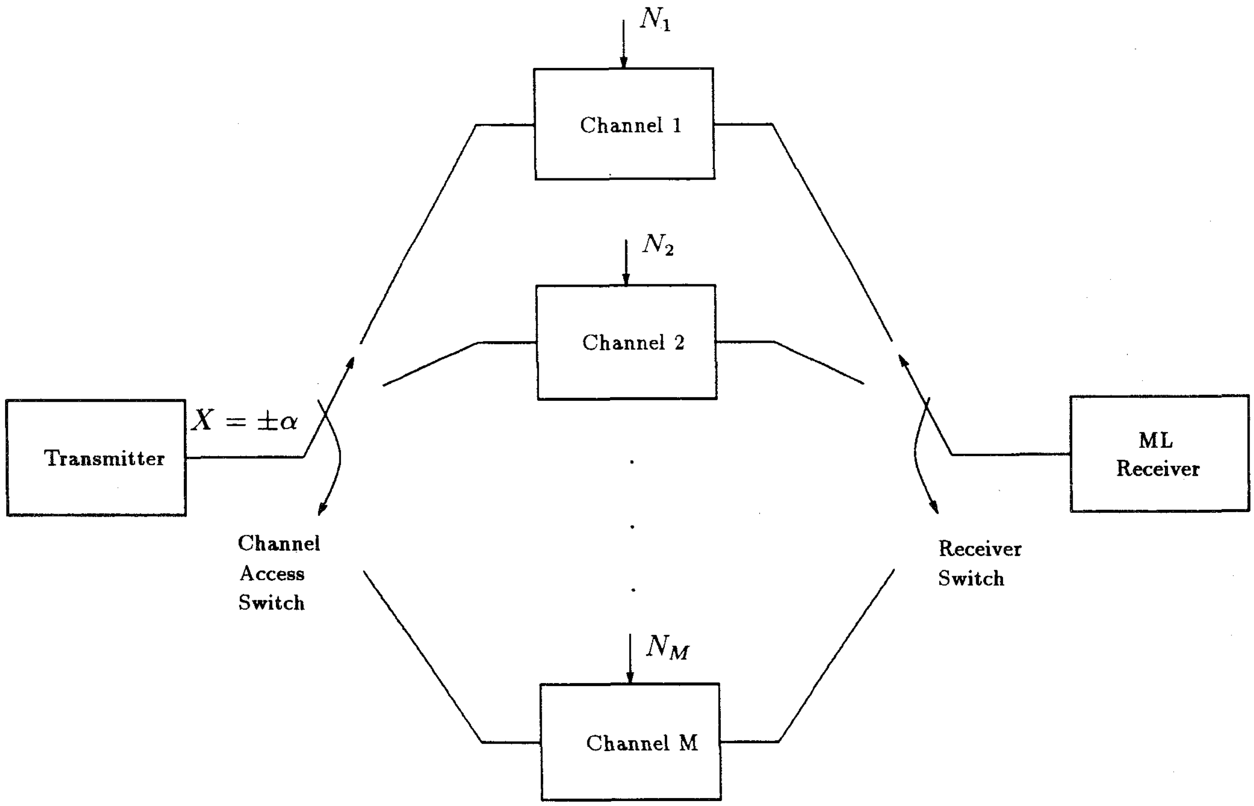

Fig. 1. The model for channel switching.

error probability for channel i. A suboptimal strategy is to use the best channel exclusively. This way one achieves

$$e_0(S) = \min\{e_i(S) : i = 1, 2, \cdots, M\} \tag{9}$$

which may be nonconvex. It should be clear that the optimum strategy will achieve

$$e^*(S) = \inf_{n, F_n} \sum_{i=1}^{n} p_i e_0(S_i) \tag{10}$$

with $n = 1, 2, \cdots$ and

$$F_n = \left\{ (p_i, S_i) : i = 1, 2, \ldots, n, 0 \le p_i \le 1, \right. \\ \left. S_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i S_i = S \right\}.$$

The following lemma, which is analogous to Lemma 1, specifies the structure of the optimal error performance.

Lemma 3: $e^*(S)$ is the largest convex function that is smaller than $e_0(S)$.

Optimum performance can be achieved by time-sharing between at most two channels and power levels [4], and yields the convex hull of individual error probability functions. For a given set of $\{e_i(S)\}$, the optimum strategy can be obtained using this observation.

In the remainder of this section we will consider the problem with two channels, as this case allows an explicit characterization of optimal channel use. We consider the case where $e_1(S)$ and $e_2(S)$ are *strictly* convex and intersect at one point S^*. That is, we assume $e_1(S) < e_2(S)$ for $S < S^*$, and $e_1(S) > e_2(S)$ for $S > S^*$. This assumption holds, for instance, when $p_1(x), p_2(x) \in \mathcal{D}_s(0)$ intersect at two points on $\Re^+$ (e.g., Gaussian and Laplace densities with the same variance).

The following lemma about the existence and uniqueness of a common tangent for $e_1(S)$ and $e_2(S)$ is proved in the Appendix.

Lemma 4: There exists a unique pair (S_1, S_2) such that the tangent to $e_1(S)$ at S_1 is also tangent to $e_2(S)$ at S_2.

The contact points S_1 and S_2 of the common tangent can be found as the unique pair that satisfies

$$e_1'(S_1) = e_2'(S_2) = \frac{e_2(S_2) - e_1(S_1)}{S_2 - S_1}$$

and they determine the optimum switching strategy as shown in the following theorem.

Theorem 4: The optimum strategy with an average power $S \in [S_1, S_2]$ uses channel 1 at power S_1 and channel 2 at power S_2. The fraction of the time channel 1 is used is $p_1 = (S_2 - S)/(S_2 - S_1)$. For $S < S_1$ (respectively, for $S > S_2$), channel 1 (respectively, channel 2) is used exclusively at the average power.

Proof: The stated strategy achieves

$$e_s(S) = \begin{cases} e_1(S), & S < S_1 \\ \frac{S_2 - S}{S_2 - S_1} e_1(S_1) + \frac{S - S_1}{S_2 - S_1} e_2(S_2), & S_1 \le S \le S_2 \\ e_2(S), & S > S_2. \end{cases}$$

It is easy to check that e_s is convex. According to Lemma 3, we must show that $e_s(S)$ is the largest convex function that is smaller than $e_0(S) = \min[e_1(S), e_2(S)]$. This clearly holds outside of $[S_1, S_2]$.

Suppose there is another convex function $f(S) \le e_0(S)$ such that $e_s(y) < f(y) \le e_0(y)$ for some $y \in [S_1, S_2]$. Then

$$e_s(y) < f(y) \le \lambda f(y_1) + (1-\lambda) f(y_2)$$
$$\le \lambda e_0(y_1) + (1-\lambda) e_0(y_2)$$

for any $0 \le \lambda \le 1$ and $\lambda y_1 + (1-\lambda)y_2 = y$. Now select $y_1 = S_1$, $y_2 = S_2$, and $\lambda = (S_2 - y)/(S_2 - S_1)$ to obtain the contradiction $e_s(y) < f(y) \le e_s(y)$. □

V. Concluding Remarks

This correspondence has obtained three properties on the convex behavior of the error probability in binary signaling with additive noise. The first result outrules the possibility of performance improvement through time-sharing when the transmitter is average-power constrained and the noise density is unimodal. The second result specifies the optimum strategy of an average-power-limited jammer, and reveals the existence of a critical power level. The third result specifies the optimum channel switching strategy in the presence of multiple additive noise channels.

The generalization of these results to nonbinary and multidimensional signaling schemes would be of interest.

Appendix

Lemma 2: Let γ^* be the inflection point of $j(\gamma)$. There exists a unique point $\gamma_c \ge \gamma^*$ such that the tangent to $j(\gamma)$ at γ_c lies above $j(\gamma)$ and passes through the origin.

Proof: The tangent to $j(\gamma)$ at point $x \ge 0$ is given by $y = j(x) + j'(x)(\gamma - x)$ and has a y-axis intercept of $h(x) = j(x) - xj'(x)$. The function $h(x)$ is continuous with derivative $h'(x) = -xj''(x)$. Since $j(\cdot)$ has a single inflection point and a finite limit, $j(\gamma)$ is concave for $\gamma > \gamma^*$ and convex for $\gamma < \gamma^*$. Thus the tangent at γ^* lies below $j(\gamma)$ for $\gamma < \gamma^*$ which implies $h(\gamma^*) \le j(0) = 0$. On the other hand, $h(\infty) = 1/2$. We have a continuous nondecreasing function $h(x)$ over (γ^*, ∞) with a negative initial value and a positive limit. Therefore, $h(x)$ must have a unique root at some point γ_c in that interval.

Since $j(\gamma)$ is concave over (γ^*, ∞), the tangent at γ_c lies above $j(\gamma)$ for $\gamma > \gamma^*$. Consider two line segments, one connecting the origin to the point $(\gamma^*, j(\gamma^*))$ and the other segment connecting the origin to the point $(\gamma_c, j(\gamma_c))$. The first line segment lies above $j(\gamma)$ for $\gamma < \gamma^*$, and has slope $j(\gamma^*)/\gamma^*$, while the second line segment has slope $j(\gamma_c)/\gamma_c$. Since

$$\frac{d}{d\gamma}\frac{j(\gamma)}{\gamma} = -\frac{h(\gamma)}{\gamma^2} \ge 0, \quad \text{for } \gamma^* \le \gamma \le \gamma_c$$

the second line segment is above the first. Hence the tangent line is above $j(\gamma)$ for $\gamma < \gamma^*$ as well. □

Lemma 4: There exists a unique pair (S_1, S_2) such that the tangent to $e_1(S)$ at S_1 is also tangent to $e_2(S)$ at S_2.

Proof: The function $\Delta e(S) = e_1(S) - e_2(S)$ has roots at $S = 0$ and at $S = S^*$. Since Δe is continuous and differentiable, it has a minimum at some point $S_{\min} \in (0, S^*)$. Also since $\lim_{S\to\infty} \Delta e(S) = 0$, Δe has a maximum at some point[3] $S_{\max} \in (S^*, \infty)$. Now consider some λ in $(S_{\min}, S^*)$. The equation $e_2'(x) = e_1'(\lambda)$ will have a root x_0 in $(\lambda, S_{\max})$ as $e_2'(\lambda) \le e_1'(\lambda) \le e_2'(S_{\max})$ and $e_2'(x)$ is continuous. Consider the function

$$h(S, \lambda) = e_2(S) - [e_1'(\lambda)(S - \lambda) + e_1(\lambda)]$$

which is the difference between $e_2(S)$ and the tangent to e_1 at λ. We have just shown that $\partial h/\partial S = e_2'(S) - e_1'(\lambda)$ has a zero at some point $x_0 \in (\lambda, S_{\max})$. Since e_2 is strictly convex $(e_2')^{-1}$ exists and $x_0 = (e_2')^{-1}(e_1'(\lambda))$ is a continuous function of λ. Then

$$g(\lambda) \triangleq h(x_0(\lambda), \lambda) = e_2(x_0) - e_1(\lambda) - e_1'(\lambda)(x_0 - \lambda)$$

is a continuous function of λ. Furthermore, $dg(\lambda)/d\lambda < 0$, since $\lambda < x_0$ and e_1 is strictly convex. Finally, when $\lambda = S_{\min}$, we have $x_0 = \lambda$, which implies

$$g(S_{\min}) = e_2(S_{\min}) - e_1(S_{\min}) > 0$$

and when $\lambda = S^*$

$$g(S^*) \le e_2(x_0) - e_2(S^*) < 0.$$

Thus the continuous decreasing function $g(\lambda)$ has a unique root $S_1 \in (S_{\min}, S^*)$. The tangent to e_1 at S_1 has a minimum vertical separation of zero from e_2 at some point $S_2 = x_0(S_1) > S^*$. This proves the existence and uniqueness of a common tangent. □

[3] It is possible that there are two or more local minima and/or local maxima. In that case, $S_{\min}$ and $S_{\max}$ are defined as those closest to S^*.

Acknowledgment

The author wishes to thank E. Telatar and R. Barry for valuable discussions about this work.

References

[1] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[2] S. Shamai and S. Verdú, "Worst-case power-constrained noise for binary-input channels," *IEEE Trans. on Information Theory*, vol. 38, pp. 1494–1511, Sept. 1992.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[4] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1968.

[5] L. C. Thomas, *Games, Theory and Applications*. Chichester, England: Ellis Horwood Ltd., 1984.

[6] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*, vol. 1. Rockville, MD: Comput. Sci. Press, 1985.

[7] R. E. Blahut, *Digital Transmission of Information*. Reading, MA: Addison-Wesley, 1990.

[8] A. J. Viterbi, "Spread spectrum communications—Myths and realities," *IEEE Commun. Mag.*, vol. 17, pp. 11–18, May 1979.

[9] A. J. Viterbi and I. M. Jacobs, "Advances in coding and modulation for noncoherent channels affected by fading, partial band, and multiple-access interference," in *Advances in Communication Systems: Theory and Applications*. New York: Academic Press, 1975, pp. 279–308.

[10] B. R. Vojcic and R. L. Pickholtz, "Performance of direct sequence spread spectrum in a fading dispersive channel with jamming," *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 561–568, May 1989.

[11] ——, "Performance of coded direct sequence spread spectrum in a fading dispersive channel with pulsed jamming," *IEEE J. Sel. Areas Commun.*, vol. 8, pp. 934–942, June 1990.

[12] R. J. McEliece and W. E. Stark, "An information theoretic study of communication in the presence of jamming," in *Proc. ICC '81*, 1981, pp. 45.3.1–45.3.5.

[13] M. V. Hegde, W. E. Stark, and D. Teneketzis, "On the capacity of channels with unknown interference," *IEEE Trans. Inform. Theory*, vol. 35, pp. 770–783, July 1989.

[14] W. E. Stark and A. A. Hassan, "Decoding algorithms with maximum error correction," in *Proc. IEEE Int. Symp. on Information Theory* (Trondheim, Norway, June 1994).