Convexity Properties of Detection Probability under Additive Gaussian Noise: Optimal Signaling and Jamming Strategies

Berkan Dulek, *Member, IEEE*, Sinan Gezici, *Senior Member, IEEE*, Orhan Arikan, *Member, IEEE*

***Abstract*—In this paper, we study the convexity properties for the problem of detecting the presence of a signal emitted from a power constrained transmitter in the presence of additive Gaussian noise under the Neyman-Pearson (NP) framework. It is proved that the detection probability corresponding to the $\alpha-$level likelihood ratio test (LRT) is either strictly concave or has two inflection points such that the function is strictly concave, strictly convex and finally strictly concave with respect to increasing values of the signal power. In addition, the analysis is extended from scalar observations to multidimensional colored Gaussian noise corrupted signals. Based on the convexity results, optimal and near-optimal time sharing strategies are proposed for average/peak power constrained transmitters and jammers. Numerical methods with global convergence are also provided to obtain the parameters for the proposed strategies.**

***Index Terms*– Detection, Neyman-Pearson (NP), Gaussian noise, convexity, stochastic signaling, jamming, time sharing, power constraint.**

I. Introduction

In coherent detection applications, despite the ubiquitous restrictions on the transmission power, there is often some flexibility in the choice of signals transmitted over the communications medium [1]. Due to crosstalk limitation between adjacent wires and frequency blocks, wired systems require that the signal power should be carefully controlled [2]. A more pronounced example from wireless systems dictates the signal power to be limited both to conserve battery power and to meet restrictions by regulatory bodies. It is well-known that the performance of optimal binary detection in Gaussian noise is improved by selecting deterministic antipodal signals along the eigenvector of the noise covariance matrix corresponding to the minimum eigenvalue [1]. Further insights are obtained by studying the convexity properties of error probability in [3] for the optimal detection of binary-valued scalar signals corrupted by additive noise under an average power constraint. It is shown that the error probability is a nonincreasing convex function of the signal power when the channel has a continuously differentiable unimodal noise probability density function (PDF) with a finite variance. This discussion is extended from binary modulations to arbitrary signal constellations in [4] by concentrating on the maximum likelihood (ML) detection over additive white Gaussian noise (AWGN) channels. The symbol error rate (SER) is shown to be always convex in signal-to-noise ratio (SNR) for 1-D and 2-D constellations, but nonconvexity in higher dimensions at low to intermediate SNRs is possible, while convexity is always guaranteed at high SNRs with an odd number of inflection points in-between. When the transmitter is average-power constrained, this result suggests the possibility of improving the error performance in high dimensional constellations through time sharing of the signal power, as opposed to the case for low dimensions (1-D and 2-D). The convexity properties of the SER with respect to jamming power (i.e., multiplicative reciprocal of SNR) are also addressed in the same study.

B. Dulek is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA, e-mail: bdulek@syr.edu.

S. Gezici and O. Arikan are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey, Tel: +90 (312) 290-3139, Fax: +90 (312) 266-4192, e-mails: {gezici,oarikan}@ee.bilkent.edu.tr.

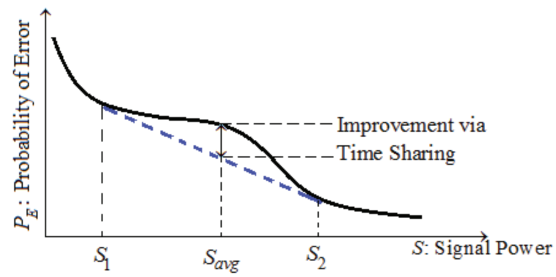

Fig. 1. Illustrative example demonstrating the benefits via time sharing between two power levels under an average power constraint.

Fig. 1 depicts how time sharing helps improve the error probability under an average power constraint via a simple illustration. Suppose that the average power constraint is denoted with $\mathrm{S_{avg}}$. It is seen that the average probability of error can be reduced by time sharing between power levels $\mathrm{S_1}$ and $\mathrm{S_2}$ with respect to the constant power transmission with $\mathrm{S_{avg}}$. More precisely, time sharing exploits the nonconvexity of the plot of error probability versus signal power. With the advent of the optimization techniques, there has been a renewed interest in designing time sharing schemes that improve/degrade (jamming problem) the error performance of communications systems operating under signal power constraints. Since performance gains in AWGN channels due to such stochastic approaches are restricted to higher dimensional constellations[1], the attempts to exploit the convexity properties of the error probability have been diverted towards channels with multimodal noise PDFs [5], [6]. Goken et al. have shown in [5] that for a given detector, the optimal signaling strategy results in a time sharing among no more than three different signal values under second and fourth moment constraints, and reported significant performance improvements over conventional signaling schemes under Gaussian mixture noise. When multiple detectors are available at the receiver of an M-ary power constrained communications system, it is stated in [6] that the optimal strategy is to time share between at most two maximum a-posteriori probability (MAP) detectors corresponding to two deterministic signal vectors.

[1] 1-D and 2-D constellations are almost universally employed in practice.

Until recently, the discussions on the benefits of stochastic signaling were severely limited to the Bayesian formulation, specifically to the error probability criterion. However, in many problems of practical interest, it is not possible to know prior probabilities or to impose specific cost structures on the decisions. In such cases, the probabilities of detection and false alarm become the main performance metrics as described in the Neyman-Pearson (NP) approach [1]. For example, in wireless sensor network applications, a transmitter can send one bit of information (using on-off keying) about the presence of an event (e.g., fire). In [7], the problem of designing the optimal signal distribution is addressed for on-off keying systems to maximize the detection probability without violating the constraints on the probability of false alarm and the average signal power. It is shown that the optimal solution can be obtained by time sharing between at most two signal vectors for the on-signal and using the corresponding NP-type likelihood ratio test (LRT) at the receiver. Although the results are general, numerical examples have been chosen from multimodal Gaussian mixture distributions to demonstrate benefits from time sharing approaches. Unfortunately even in that case, finding the optimal signal set to maximize the detection probability is a computationally cumbersome task necessitating the use of global optimization techniques [7].

In this paper, we report an interesting and obviously overlooked fact for the problem of detecting the presence of a signal emitted from a power constrained transmitter operating over an additive Gaussian noise channel within the NP framework. Contrary to the error probability criterion [4], it is shown that for false alarm rates smaller than $Q(2)$, remarkable improvements in detection probability can be attained even in low dimensions by optimally distributing the fixed average power between two levels ($Q(\cdot)$ denotes the $Q-$function). More specifically, we study analytically the convexity properties of determining the presence of a power-limited signal immersed in additive Gaussian noise. It is proved that the detection probability corresponding to the $\alpha-$level LRT is either concave for $\alpha \geq Q(2)$ or has two inflection points such that the function is strictly concave, strictly convex and finally strictly concave with respect to increasing values of the signal power for $\alpha < Q(2)$. Numerical methods with global convergence are provided to determine the regions over which time sharing enhances the detection performance over deterministic signaling at the average power level. In addition, the analysis is extended from scalar observations to multidimensional colored Gaussian noise corrupted signals. Based on the convexity results, optimal and near-optimal time sharing strategies are proposed for average/peak power constrained transmitters. For almost all practical applications, the required false alarm probability values are much smaller than $Q(2) \approx 0.02275$. As a consequence, time sharing can facilitate improved detection performance whenever the average power limitations are in the designated regions. Finally, the dual problem is considered from the perspective of a Gaussian jammer to decrease the detection probability via time sharing. It is shown that the optimal strategy results in on-off jamming when the average noise power is below some critical value, a fact previously noted for spread spectrum communications systems [8].

II. Problem Formulation

Consider the problem of detecting the presence of a target signal, where the receiver needs to decide between the two hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$ based on a real-valued scalar observation Y acquired over an AWGN channel.

$$\mathcal{H}_0 : \; Y = \sigma N \quad , \quad \mathcal{H}_1 \; : \; Y = \sqrt{S} + \sigma N \tag{1}$$

Here, $N \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable with zero mean and unit variance, $\sigma > 0$ is the noise standard deviation at the receiver, $\sqrt{S}$ represents the transmitted signal for the alternative hypothesis $\mathcal{H}_1$, and $S > 0$ is the corresponding signal power. The additive noise N is statistically independent of the signal $\sqrt{S}$. The scalar channel model in (1) provides an abstraction for a continuous-time system that passes the received signal through a correlator (matched filter) and samples it once per symbol interval, thereby capturing the effects of modulator, additive noise channel and receiver front-end processing. In addition, although the above model is in the form of a simple additive noise channel, it may be sufficient to incorporate various effects such as thermal noise, multiple-access interference, and jamming [3].

It is well-known that the NP detector gives the most powerful α-level test of $\mathcal{H}_0$ versus $\mathcal{H}_1$ [1]. In other words, when the aim is to maximize the probability of detection such that the probability of false alarm does not exceed a predetermined value α, the NP detector is the optimal choice and takes the following form of an LRT for continuous PDFs:

$$\delta_{NP}(y) = \begin{cases} 1\,, & \text{if} \;\; p_1(y) \geq \eta\, p_0(y) \\ 0\,, & \text{if} \;\; p_1(y) < \eta\, p_0(y) \end{cases} \tag{2}$$

where the threshold $\eta \geq 0$ is chosen such that the probability of false alarm satisfies $\mathrm{P_{FA}} = \mathrm{P}_0\left(p_1(y) \geq \eta\, p_0(y)\right) = \alpha$, with subscript 0 denoting that the probability is calculated conditioned on the null hypothesis $\mathcal{H}_0$. Then, the NP decision rule is the optimal one among all α-level decision rules, i.e., $\mathrm{P_D} = \mathrm{P}_1\left(p_1(y) \geq \eta\, p_0(y)\right)$ is maximized, where the probability is calculated under the condition that the alternative hypothesis $\mathcal{H}_1$ is true.

The hypothesis pair in (1) can be restated in terms of the distributions on the observation space as $\mathcal{H}_0 \, : \, Y \sim \mathcal{N}(0,\sigma^2)$ and $\mathcal{H}_1 \, : Y \sim \mathcal{N}(\sqrt{S},\sigma^2)$. The likelihood ratio for (1) is then given by $L(y) = p_1(y)/p_0(y) = \exp\left\{\sqrt{S}/\sigma^2\left(y - \sqrt{S}/2\right)\right\}$. Since $S > 0$, the likelihood ratio $L(y)$ is a strictly increasing function of the observation y. Therefore, comparing $L(y)$ to the threshold η is equivalent to comparing y to another threshold $\eta' = L^{-1}(\eta)$, where L^{-1} is the inverse function of L. Then, the probability of false alarm is expressed as $\mathrm{P_{FA}} = \mathrm{P}_0\left(L(Y) \geq \eta\right) = \mathrm{P}_0\left(Y \geq \eta'\right) = Q\left(\eta'/\sigma\right)$, where $Q-$function is the tail probability of the standard Gaussian distribution, i.e., $Q(x) = (1/\sqrt{2\pi}\,)\int_x^\infty \mathrm{e}^{-t^2/2} dt$. It is noted that any value of false alarm probability α can be attained by choosing the threshold $\eta' = \sigma Q^{-1}(\alpha)$, where Q^{-1} is the inverse $Q-$function. Then, for fixed S, the optimal $\alpha-$level NP decision rule employed at the receiver is given by

$$\delta_{NP}(y) = \begin{cases} 1\,, & \text{if } y \geq \sigma Q^{-1}(\alpha) \\ 0\,, & \text{if } y < \sigma Q^{-1}(\alpha) \end{cases} \tag{3}$$

which also possesses the constant false alarm rate (CFAR) property [1]. Let $\gamma \triangleq S/\sigma^2$ denote the normalized signal power at the receiver. Then, the detection probability achieved by δ_{NP} is obtained as

$$\mathrm{P_D}(\gamma) = \mathrm{P}_1\left(Y \geq \sigma Q^{-1}(\alpha)\right) = Q\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right). \qquad (4)$$

For fixed α, the relationship between the detection probability and γ is known as the power function of the test in radar terminology [1].

We will first discuss the convexity properties of the detection probability with respect to the signal power for the NP test given in (3). This is motivated by the possibility of enhancing the detection performance via time sharing between two signal power levels while satisfying an average power constraint [3], [4], [9]. In the absence of fading, the average received power is a deterministically scaled version of the transmitted power for non-varying AWGN channels. Hence, any constraint on the transmitted power can be related to one on the received power and consecutively to one in the normalized form, and vice versa. In addition to the average power constraint, a hard limit on the peak transmitted power can be imposed as well in accordance with practical considerations.

III. Convexity Properties in Signal Power

A. Convexity/Concavity Results

In the following analysis, the endpoints are excluded from the set of feasible false alarm probabilities. Specifically, α is confined in the interval $(0, 1)$ excluding the trivial cases of $\alpha \in \{0, 1\}$. We first note the limits of the detection probability, i.e., $\lim_{\gamma\to 0} \mathrm{P_D}(\gamma) = \alpha$ and $\lim_{\gamma\to\infty} \mathrm{P_D}(\gamma) = 1$. Differentiating with respect to γ yields $\mathrm{P}'_{\mathrm{D}}(\gamma) = (2\sqrt{2\pi\gamma})^{-1} \exp\left\{-\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)^2/2\right\}$, which is positive $\forall\,\gamma > 0$ indicating that $\mathrm{P_D}(\gamma)$ is a strictly increasing function of γ. Similarly, the limits for the first derivative is given as $\lim_{\gamma\to 0} \mathrm{P}'_{\mathrm{D}}(\gamma) = \infty$ and $\lim_{\gamma\to\infty} \mathrm{P}'_{\mathrm{D}}(\gamma) = 0$.

Proposition 1: For $\alpha \in [Q(2), 1)$, $\mathrm{P_D}(\gamma)$ is a monotonically increasing and strictly concave function of $\gamma \in (0, \infty)$. For $\alpha \in (0, Q(2))$, $\mathrm{P_D}(\gamma)$ is a monotonically increasing function with two inflection points $\gamma_1 < \gamma_2$ such that $\mathrm{P_D}(\gamma)$ is strictly concave for $\gamma \in (0, \gamma_1)$, strictly convex for $\gamma \in (\gamma_1, \gamma_2)$, and strictly concave for $\gamma \in (\gamma_2, \infty)$.

Proof: It suffices to consider the second derivative of the detection probability with respect to γ, i.e.,

$$\mathrm{P}''_{\mathrm{D}}(\gamma) = \frac{1}{4\sqrt{2\pi}\,\gamma} \exp\left\{-\frac{\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)^2}{2}\right\} \times \left(Q^{-1}(\alpha) - \sqrt{\gamma} - \frac{1}{\sqrt{\gamma}}\right). \qquad (5)$$

Since the first two terms in (5) are positive $\forall\,\gamma > 0$, the sign of the second derivative is determined by the third term, i.e., $(Q^{-1}(\alpha) - \sqrt{\gamma} - 1/\sqrt{\gamma})$. First, it is noted that for $\alpha \geq Q(0) = 0.5$, we have $Q^{-1}(\alpha) \leq 0$ which implies $\mathrm{P}''_{\mathrm{D}}(\gamma) < 0$ for all $\gamma > 0$ and the detection probability is strictly concave. Next, let $x \triangleq \sqrt{\gamma}$. The third term in (5) has the reversed sign of $f(x) = x^2 - Q^{-1}(\alpha)x + 1$ for $x > 0$. The sign of quadratic polynomial $f(x)$ can be determined from its discriminant, which is given by $\Delta = (Q^{-1}(\alpha))^2 - 4$. When $\alpha \in (Q(2), Q(-2))$, the discriminant is negative $\Delta < 0$, and we have $f(x) > 0\ \forall\, x$. Both $\alpha \geq Q(0)$ and $\alpha \in (Q(2), Q(-2))$ imply that $\mathrm{P}''_{\mathrm{D}}(\gamma) < 0$. Thus , it is concluded that $\mathrm{P_D}(\gamma)$ is strictly concave

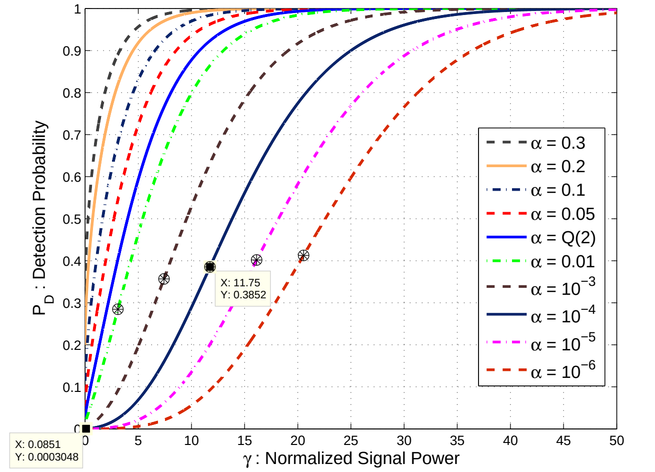

Fig. 2. Detection probability of the NP decision rule in (3) is plotted versus γ for various values of the false alarm probability α. As an example, when $\alpha = 10^{-4}$, the inflection points are located at $\gamma_1 \approx 0.0851$ and $\gamma_2 \approx 11.7459$ with $\mathrm{P_D}(\gamma_1) \approx 0.0003$ and $\mathrm{P_D}(\gamma_2) \approx 0.3852$. The second inflection point (γ_2) is also marked on each curve for $\alpha > Q(2)$.

for $\alpha > Q(2) \approx 0.02275013$. For $\alpha < Q(2)$, $f(x)$ has two distinct roots corresponding to the inflection points of $\mathrm{P_D}(\gamma)$, which are given as

$$\gamma_1 = 0.25\left(Q^{-1}(\alpha) - \sqrt{(Q^{-1}(\alpha))^2 - 4}\right)^2$$
$$\gamma_2 = 0.25\left(Q^{-1}(\alpha) + \sqrt{(Q^{-1}(\alpha))^2 - 4}\right)^2 \qquad (6)$$

suggesting that $\mathrm{P_D}(\gamma)$ is strictly concave for $\gamma \in (0, \gamma_1) \cup (\gamma_2, \infty)$ and strictly convex for $\gamma \in (\gamma_1, \gamma_2)$. □

Fig. 2 depicts the detection probability of the NP decision rule in (3) versus γ for various values of the false alarm probability α. As expected, $\mathrm{P_D}(\gamma)$ is strictly concave for $\alpha \in [Q(2), 1)$, and consists of strictly concave, strictly convex and finally strictly concave intervals for $\alpha \in (0, Q(2))$. For the latter case, even though its existence is guaranteed, the effect of the first inflection point is far less obvious than the second inflection point. This can be attributed to the fact that for small values of α, $\gamma_1 \approx 0$ and $\mathrm{P_D}(\gamma_1) \approx \alpha$ whereas $\gamma_2 \approx (Q^{-1}(\alpha))^2$ and $\mathrm{P_D}(\gamma_2) \approx 0.5$, where the approximations are obtained using the first order Taylor series expansion.

B. Optimal Signaling

The concavity of detection probability for $\alpha \in [Q(2), 1)$ stated in Proposition 1 indicates that the detection performance of an average power-limited transmitter cannot be improved by time sharing between different power levels. This follows from Jensen's inequality since the detection probability achieved via time sharing, which is the convex combination of detection probabilities corresponding to different power levels, is always smaller than the detection probability when transmitting at a fixed power that is equal to the same convex combination of the power levels. Fortunately, the range of false alarm probabilities facilitating improved detection performance, $\alpha \in (0, Q(2))$, have higher practical significance. In order to obtain the optimal time sharing strategy, we first present the following lemma which is proved in the Appendix.

Lemma 1: Let $\alpha < Q(2)$, and γ_1 and γ_2 be the inflection points of $\mathrm{P_D}(\gamma)$ as given in (6). There exist unique points $\gamma_{C1} \in (0, \gamma_1]$ and $\gamma_{C2} \geq \gamma_2$ such that the tangent to $\mathrm{P_D}(\gamma)$ at γ_{C1} is also tangent at γ_{C2} and this tangent lies above $\mathrm{P_D}(\gamma)$ for all $\gamma > 0$.

Using a similar analysis to that in the proof of Lemma 1, we can also obtain the following lemma.

Lemma 2: Let $\alpha < Q(2)$, and γ_1 and γ_2 be the inflection points of $\mathrm{P_D}(\gamma)$. Suppose also that γ_{C1} and γ_{C2} are the contact points of the tangent line as described in Lemma 1. Given a point $\hat{\gamma} \in [\gamma_1, \gamma_{C2}]$, there exists a unique point $\gamma_C(\hat{\gamma}) \in [\gamma_{C1}, \gamma_1]$ such that the tangent at $\gamma_C(\hat{\gamma})$ passes through the point $(\hat{\gamma}, \mathrm{P_D}(\hat{\gamma}))$ and lies above $\mathrm{P_D}(\gamma)$ for all $\gamma \in (0, \hat{\gamma})$.[2]

Based on Lemma 1 and Lemma 2, we state the optimal signaling strategy for the communications system in (1) operating under peak power constraint Γ_{peak} and average power constraint Γ_{avg} ($\Gamma_{\mathrm{avg}} \leq \Gamma_{\mathrm{peak}}$).

Proposition 2: Let $\alpha < Q(2)$. For $\Gamma_{\mathrm{avg}} \leq \gamma_{C1}$ or $\Gamma_{\mathrm{avg}} \geq \gamma_{C2}$ or $\Gamma_{\mathrm{peak}} \leq \gamma_1$, the best strategy is to exclusively transmit at the average power Γ_{avg}, i.e., time sharing does not help. When $\Gamma_{\mathrm{avg}} \in (\gamma_{C1}, \gamma_{C2})$ and $\gamma_{C2} \leq \Gamma_{\mathrm{peak}}$, the optimal strategy is to time share between powers γ_{C1} and γ_{C2} with the fraction of time $(\gamma_{C2} - \Gamma_{\mathrm{avg}})/(\gamma_{C2} - \gamma_{C1})$ allocated to the power γ_{C1}. On the contrary if $\Gamma_{\mathrm{avg}} \in [\gamma_C(\Gamma_{\mathrm{peak}}), \Gamma_{\mathrm{peak}}]$ while $\Gamma_{\mathrm{peak}} \in (\gamma_1, \gamma_{C2})$, the optimal strategy is to time share between powers $\gamma_C(\Gamma_{\mathrm{peak}})$ and the peak power Γ_{peak} with the fraction of time $(\Gamma_{\mathrm{peak}} - \Gamma_{\mathrm{avg}})/(\Gamma_{\mathrm{peak}} - \gamma_C(\Gamma_{\mathrm{peak}}))$ allocated to the power $\gamma_C(\Gamma_{\mathrm{peak}})$. Consequently, if $\Gamma_{\mathrm{avg}} < \gamma_C(\Gamma_{\mathrm{peak}})$ while $\Gamma_{\mathrm{peak}} \in (\gamma_1, \gamma_{C2})$, transmitting continuously at Γ_{avg} is the optimal strategy.[3]

Proof: We state the proof in the absence of a peak power constraint. Let $\lambda_T \triangleq (\mathrm{P_D}(\gamma_{C2}) - \mathrm{P_D}(\gamma_{C1}))\,/\,(\gamma_{C2} - \gamma_{C1})$. For an average power γ, the proposed strategy achieves

$$\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma) = \begin{cases} \mathrm{P_D}(\gamma) & \text{if } \gamma \in (0, \gamma_{C1}) \cup (\gamma_{C2}, \infty) \\ \mathrm{P_D}(\gamma_{C1}) + \lambda_T\,(\gamma - \gamma_{C1}) & \text{if } \gamma \in [\gamma_{C1}, \gamma_{C2}] \end{cases} \tag{7}$$

It is easy to see that $\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)$ is concave. Next, we need to show that the detection probability cannot be increased any further by time sharing between different power levels. More precisely, $\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)$ is the smallest concave function that is larger than $\mathrm{P_D}(\gamma)$ [3]. For $\gamma \in (0, \gamma_{C1}) \cup (\gamma_{C2}, \infty)$, this clearly holds. For $\gamma \in [\gamma_{C1}, \gamma_{C2}]$, the proof is via contradiction. Suppose that there exists another concave function $g(\gamma)$ greater than $\mathrm{P_D}(\gamma)$ with the property $\tilde{\mathrm{P}}_{\mathrm{D}}(x) > g(x) \geq \mathrm{P_D}(x)$ for some $x \in [\gamma_{C1}, \gamma_{C2}]$. Due to concavity of $g(x)$, we have $\tilde{\mathrm{P}}_{\mathrm{D}}(x) > g(x) \geq \theta g(x_1) + (1-\theta) g(x_2) \geq \theta \mathrm{P_D}(x_1) + (1-\theta)\mathrm{P_D}(x_2)$ for any $\theta \in [0,1]$ and $x = \theta x_1 + (1-\theta)x_2$. Now let $x_1 = \gamma_{C1}$, $x_2 = \gamma_{C2}$, and $\theta = (\gamma_{C2} - x)\,/\,(\gamma_{C2} - \gamma_{C1})$. Then, $\tilde{\mathrm{P}}_{\mathrm{D}}(x) > g(x) \geq \tilde{\mathrm{P}}_{\mathrm{D}}(x)$, which is a contradiction. This completes the proof. The proofs for the proposed time sharing strategies that are detailed according to the various relations among $\Gamma_{\mathrm{peak}}, \Gamma_{\mathrm{avg}}, \gamma_1, \gamma_2, \gamma_{C1}$ and γ_{C2} can be obtained similarly. □

It should be noted that the transmitter requires the knowledge of the noise variance at the receiver in order to employ the optimal time sharing strategy. If we do not pay attention to the peak power constraint for a second, these results indicate that very weak and strong transmitters should operate continuously at their average power while transmitters with moderate power can benefit significantly from time sharing strategies.

The critical points γ_{C1} and γ_{C2} can be obtained as the unique pair that satisfies $\mathrm{P}'_{\mathrm{D}}(\gamma_{C1}) = \mathrm{P}'_{\mathrm{D}}(\gamma_{C2}) = (\mathrm{P_D}(\gamma_{C2}) - \mathrm{P_D}(\gamma_{C1}))\,/\,(\gamma_{C2} - \gamma_{C1})$, which can be solved numerically by plugging in the corresponding expressions. Since the simultaneous solution of these equality constraints can be difficult due to terms involving exponentials and $Q-$functions, we propose two approaches to obtain the optimal signaling strategy. The first is to solve the following nonconvex optimization problem:

$$\begin{aligned} \max_{\lambda, \gamma_{C1}, \gamma_{C2}} \quad & \lambda\, Q\left(Q^{-1}(\alpha) - \sqrt{\gamma_{C1}}\right) + (1-\lambda) Q\left(Q^{-1}(\alpha) - \sqrt{\gamma_{C2}}\right) \\ \text{s.t.} \quad & \lambda\, \gamma_{C1} + (1-\lambda)\gamma_{C2} \leq \Gamma_{\mathrm{avg}} \end{aligned} \tag{8}$$

where $\gamma_{C1} \in (0, \gamma_1]$, $\gamma_{C2} \in [\gamma_2, \Gamma_{\mathrm{peak}}]$, and $\lambda \in [0,1]$ denotes the fraction of time power γ_{C1} is used assuming $\Gamma_{\mathrm{peak}} \geq \gamma_{C2}$ and $\Gamma_{\mathrm{avg}} \in [\gamma_{C1}, \gamma_{C2}]$. A local solver can be employed using multiple start points that are uniformly distributed within the bounds. The global optimum can then be selected among those local maxima by returning the one with the maximum score. In our trials, we observe that close to optimal solutions can be obtained using as few as 10 start points from each interval without compromising the computational efficiency.

A much more effective numerical method to obtain the unique tangent points $\gamma_{C1} \in (0, \gamma_1]$ and $\gamma_{C2} \geq \gamma_2$ is presented next. Based on a bisection search, this method is guaranteed to converge to the exact values for γ_{C1} and γ_{C2} with desired accuracy. More explicitly, we propose the following algorithm.

Algorithm 1

$\lambda_{min} = \mathrm{P}'_{\mathrm{D}}(\gamma_1),\ \lambda_{max} = \mathrm{P}'_{\mathrm{D}}(\gamma_2)$
$\gamma_{min,1} = 0,\ \gamma_{max,1} = \gamma_1$
$\gamma_{min,2} = \gamma_2,\ \gamma_{max,2} = \infty$
do
  $\lambda = (\lambda_{max} + \lambda_{min})\,/2$
  $\gamma_{X1} = \underset{\gamma \in (\gamma_{min,1}, \gamma_{max,1})}{\operatorname{argmax}} \mathrm{P_D}(\gamma) - \lambda\gamma$
  $\gamma_{X2} = \underset{\gamma \in (\gamma_{min,2}, \gamma_{max,2})}{\operatorname{argmax}} \mathrm{P_D}(\gamma) - \lambda\gamma$
  if $\mathrm{P_D}(\gamma_{X1}) - \lambda\gamma_{X1} > \mathrm{P_D}(\gamma_{X2}) - \lambda\gamma_{X2}$,
  then $\lambda_{max} = \lambda,\ \gamma_{min,1} = \gamma_{X1},\ \gamma_{min,2} = \gamma_{X2}$
  else $\lambda_{min} = \lambda,\ \gamma_{max,1} = \gamma_{X1},\ \gamma_{max,2} = \gamma_{X2}$
while $|(\mathrm{P_D}(\gamma_{X1}) - \lambda\gamma_{X1}) - (\mathrm{P_D}(\gamma_{X2}) - \lambda\gamma_{X2})| > \epsilon$

To see that the tangent points γ_{C1} and γ_{C2} can be obtained via the proposed algorithm, a few observations are noted first. The slope of $\mathrm{P_D}(\gamma)$ strictly decreases in the interval $[\gamma_{C1}, \gamma_1]$, strictly increases in the interval $[\gamma_1, \gamma_2]$, and then again strictly decreases in the interval $[\gamma_2, \gamma_{C2}]$. Consequently, we have $\mathrm{P}'_{\mathrm{D}}(\gamma_1) < \mathrm{P}'_{\mathrm{D}}(\gamma_{C1}) = \mathrm{P}'_{\mathrm{D}}(\gamma_{C2}) < \mathrm{P}'_{\mathrm{D}}(\gamma_2)$. Using the analytical expressions derived for γ_1, γ_2 and $\mathrm{P}'_{\mathrm{D}}(\gamma)$, the computations of $\mathrm{P}'_{\mathrm{D}}(\gamma_1)$ and $\mathrm{P}'_{\mathrm{D}}(\gamma_2)$ are straightforward. Hence, initial lower and upper bounds are obtained for the slope of $\mathrm{P_D}(\gamma)$ at the tangent points γ_{C1} and γ_{C2}. These are denoted with λ_{min} and λ_{max} at the beginning of the proposed algorithm, respectively.

[2] The dependence of tangent point γ_C to $\hat{\gamma}$ is explicitly emphasized by writing it as a function, i.e., $\gamma_C(\hat{\gamma})$.

[3] The cases of $\Gamma_{\mathrm{avg}} \leq \gamma_{C1}$ and $\Gamma_{\mathrm{peak}} \leq \gamma_1$ can be practically uninteresting since they result in very low detection probabilities.

Let $\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)$ and λ_T be as defined in (7). It is noted that $\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)$ represents the upper boundary of the convex hull of $\mathrm{P}_{\mathrm{D}}(\gamma)$. Now consider the function $\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma$ for $\lambda>0$. Since $\mathrm{P}_{\mathrm{D}}(\gamma)\leq\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)$ for all $\gamma>0$, we have $\max_{\gamma>0}\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma\leq\max_{\gamma>0}\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)-\lambda\gamma$. The maximum of the right-hand side occurs at $\tilde{\mathrm{P}}_{\mathrm{D}}^{'}(\gamma)=\lambda$, for which a unique solution exists for all positive $\lambda\neq\lambda_T$. This is because $\tilde{\mathrm{P}}_{\mathrm{D}}^{'}(\gamma)=\mathrm{P}_{\mathrm{D}}^{'}(\gamma)$ over the intervals $(0,\gamma_{C1}]$ and $[\gamma_{C2},\infty)$, where $\mathrm{P}_{\mathrm{D}}^{'}(\gamma)$ is strictly decreasing and continuous with $\lim_{\gamma\to0}\mathrm{P}_{\mathrm{D}}^{'}(\gamma)=\infty$, $\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{C1})=\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{C2})=\lambda_T$ and $\lim_{\gamma\to\infty}\mathrm{P}_{\mathrm{D}}^{'}(\gamma)=0$. Hence, we have $\max_{\gamma>0}\tilde{\mathrm{P}}_{\mathrm{D}}(\gamma)-\lambda\gamma=\max_{\gamma>0}\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma$ for all $\lambda>0$. More explicitly, by defining $\hat{\gamma}(\lambda)\triangleq\mathrm{argmax}_{\gamma\in(0,\infty)}\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma$, it is seen that $\hat{\gamma}(\lambda)$ is a decreasing function of λ with $\hat{\gamma}(\lambda)\in(0,\gamma_{C1})$ for $\lambda>\lambda_T$ and $\hat{\gamma}(\lambda)\in(\gamma_{C2},\infty)$ for $0<\lambda<\lambda_T$.

These observations are exploited in Algorithm 1 as follows. Since $\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma$ is *strictly concave* over the intervals $(0,\gamma_1]$ and $[\gamma_2,\infty)$, γ_{X1} and γ_{X2} can be computed efficiently at each iteration by means of convex optimization methods. Furthermore, the bounds denoted with $\gamma_{min,i},\gamma_{max,i}$ get tighter with each iteration for $i=1,2$. Suppose that $\lambda>\lambda_T$ at the first iteration. Then, the maximum is attained within the interval $(0,\gamma_{C1})$ and $\mathrm{P}_{\mathrm{D}}(\gamma_{X1})-\lambda\gamma_{X1}>\mathrm{P}_{\mathrm{D}}(\gamma_{X2})-\lambda\gamma_{X2}$ is satisfied. Since $\lambda_T<\lambda$, all values greater than the current value of λ are discarded by setting $\lambda_{max}=\lambda$. Likewise, since $\gamma_{C1}>\gamma_{X1}$ and $\gamma_{C2}>\gamma_{X2}$, all the values smaller than the current values of γ_{X1} and γ_{X2} are discarded from the search intervals for the next values of γ_{X1} and γ_{X2}, respectively. If $\lambda<\lambda_T$ at the first iteration, the maximum is attained within the interval (γ_{C2},∞) and $\mathrm{P}_{\mathrm{D}}(\gamma_{X1})-\lambda\gamma_{X1}<\mathrm{P}_{\mathrm{D}}(\gamma_{X2})-\lambda\gamma_{X2}$ is satisfied. In this case, we have $\lambda_T>\lambda$ and all values smaller than the current value of λ are discarded by setting $\lambda_{min}=\lambda$. Likewise, since $\gamma_{C1}<\gamma_{X1}$ and $\gamma_{C2}<\gamma_{X2}$, all the values greater than the current values of γ_{X1} and γ_{X2} are discarded from the search intervals for the next values of γ_{X1} and γ_{X2}, respectively. At each iteration, either λ_{min} increases towards λ_T or λ_{max} decreases towards λ_T, and $\lambda_{max}\geq\lambda_T\geq\lambda_{min}$ is assured. Thus, λ converges to λ_T. At convergence, we have $\gamma_{X1}=\mathrm{argmax}_{\gamma\in(\gamma_{min,1},\gamma_{max,1})}\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda_T\gamma=\gamma_{C1}$ and $\gamma_{X2}=\mathrm{argmax}_{\gamma\in(\gamma_{min,2},\gamma_{max,2})}\mathrm{P}_{\mathrm{D}}(\gamma)-\lambda_T\gamma=\gamma_{C2}$. In practice, a sufficiently small value is selected for ϵ to control the accuracy of the solution at convergence.

Proposition 2 requires also the knowledge of $\gamma_C(\Gamma_{\mathrm{peak}})$ for the optimal signaling strategy in the case of $\Gamma_{\mathrm{peak}}\in(\gamma_1,\gamma_{C2})$, where $\gamma_C(\Gamma_{\mathrm{peak}})$ is as defined in Lemma 2. A similar bisection search can be used to find $\gamma_C(\Gamma_{\mathrm{peak}})$ after γ_{C1} and γ_{C2} are obtained via Algorithm 1. This method is described in Algorithm 2, the proof of which can be stated similarly.

Algorithm 2

$\lambda_{min}=\mathrm{P}_{\mathrm{D}}^{'}(\gamma_1),\ \lambda_{max}=\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{C1})$
$\gamma_{min}=\gamma_{C1},\ \gamma_{max}=\gamma_1$
do
  $\lambda=(\lambda_{max}+\lambda_{min})/2$
  $\gamma_X=\underset{\gamma\in(\gamma_{min},\gamma_{max})}{\mathrm{argmax}}\ \mathrm{P}_{\mathrm{D}}(\gamma)-\lambda\gamma$
  if $\mathrm{P}_{\mathrm{D}}(\gamma_X)+\mathrm{P}_{\mathrm{D}}^{'}(\gamma_X)(\Gamma_{\mathrm{peak}}-\gamma_X)>\mathrm{P}_{\mathrm{D}}(\Gamma_{\mathrm{peak}})$,
  then $\lambda_{max}=\lambda,\ \gamma_{min}=\gamma_X$
  else $\lambda_{min}=\lambda,\ \gamma_{max}=\gamma_X$
while $\left|\mathrm{P}_{\mathrm{D}}(\gamma_X)+\mathrm{P}_{\mathrm{D}}^{'}(\gamma_X)(\Gamma_{\mathrm{peak}}-\gamma_X)-\mathrm{P}_{\mathrm{D}}(\Gamma_{\mathrm{peak}})\right|>\epsilon$

As an example, for $\alpha=10^{-4}$, $\Gamma_{\mathrm{avg}}=5$, and $\Gamma_{\mathrm{peak}}=20$, the optimal strategy can achieve a detection probability of 0.1946 by employing power $\gamma_{C1}=2.69\times10^{-5}$ with probability 0.7307 and power $\gamma_{C2}=18.5664$ with probability 0.2693, whereas by exclusively transmitting at the average power, the detection probability remains at 0.0690. If the peak power constraint is lowered to $\Gamma_{\mathrm{peak}}=10$, the optimal strategy can still increase the detection probability to 0.1445 by time sharing between $\gamma_C=4.99\times10^{-5}$ and peak power $\Gamma_{\mathrm{peak}}=10$ with approximately equal fractions as suggested by the solution of $\mathrm{P}_{\mathrm{D}}^{'}(\gamma_C)=(\mathrm{P}_{\mathrm{D}}(\Gamma_{\mathrm{peak}})-\mathrm{P}_{\mathrm{D}}(\gamma_C))/(\Gamma_{\mathrm{peak}}-\gamma_C)$. Finally, it should be emphasized that the detection probability can be improved even further by designing the optimal signaling scheme jointly with the detector employed at the receiver as discussed in [7]. However, in that case we need to sacrifice from the simplistic structure of the threshold detector which is also easier to update if the channel statistics change slowly over time.

C. Near-optimal Strategy

It should be noted that Algorithm 1 requires the solution of two convex optimization problems at each iteration to obtain the critical points γ_{C1} and γ_{C2}, that are needed to describe the optimal signaling strategy. Moreover, $\gamma_C(\Gamma_{\mathrm{peak}})$ should also be determined using Algorithm 2 whenever $\Gamma_{\mathrm{peak}}\in(\gamma_1,\gamma_{C2})$. In the following, it is shown that near-optimal performance can be achieved with computational complexity comparable to only that of Algorithm 2.

We recall from the previous discussion that for small values of the false alarm probability, the first inflection point γ_1 gets close to zero. It is also stated above that the value of $\mathrm{P}_{\mathrm{D}}(\gamma_1)$ equals approximately to α in that case. Since the critical points γ_C and γ_{C1} are located inside the interval $(0,\gamma_1]$, they get close to zero as well while the corresponding detection probabilities approach α. Also evident from the example above, this observation gives clues of a suboptimal approach. We make a simplifying assumption and suppose that $\mathrm{P}_{\mathrm{D}}(\gamma)$ is convex over the interval $(0,\gamma_2)$. Using arguments similar to those in the Appendix, it is then possible to show that there exists a unique point $\gamma_{\mathrm{on}}\geq\gamma_2$ such that the tangent to $\mathrm{P}_{\mathrm{D}}(\gamma)$ at γ_{on} passes through the point $(0,\alpha)$. This observation leads to the following near-optimal strategy in the case of strict false alarm requirements.

Near-optimal strategy: Let $\alpha<Q(2)$. A suboptimal strategy with reasonable performance is to switch between powers 0 and γ_{on} with the fraction of on-power time $\Gamma_{\mathrm{avg}}/\gamma_{\mathrm{on}}$ when $\Gamma_{\mathrm{avg}}<\gamma_{\mathrm{on}}<\Gamma_{\mathrm{peak}}$. For $\gamma_{\mathrm{on}}\geq\Gamma_{\mathrm{peak}}$, the proposed suboptimal strategy time shares between powers 0 and Γ_{peak} with the fraction of on-power time $\Gamma_{\mathrm{avg}}/\Gamma_{\mathrm{peak}}$. For $\Gamma_{\mathrm{avg}}>\gamma_{\mathrm{on}}$, the transmission is conducted exclusively at the average power.

γ_{on} can be obtained from $\mathrm{P}_{\mathrm{D}}(\gamma_{\mathrm{on}})-\gamma_{\mathrm{on}}\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{\mathrm{on}})=\alpha$. More explicitly, we need to solve for $\hat{x}$ such that

$$\hat{x}=Q^{-1}\left(\frac{Q^{-1}(\alpha)-\hat{x}}{2\sqrt{2\pi}}\exp\left\{-\frac{\hat{x}^2}{2}\right\}+\alpha\right)\tag{9}$$

and the contact point can be obtained by substituting $\gamma_{\mathrm{on}}=\left(Q^{-1}(\alpha)-\hat{x}\right)^2$. The form of the equation in (9) suggests that a fixed point iteration can be employed to obtain the solution [10]. However,

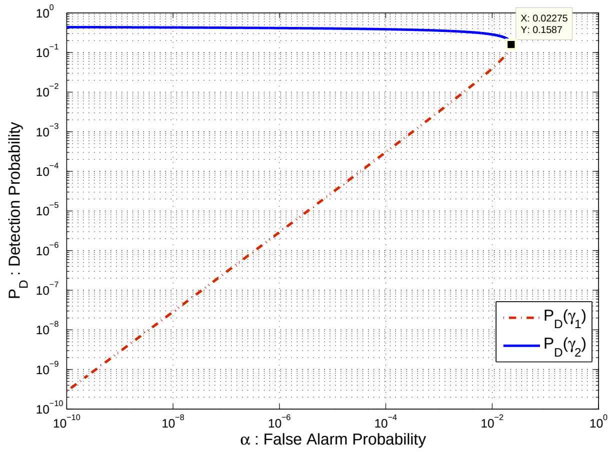

Fig. 3. Detection probability of the NP decision rule in (3) is evaluated at the inflection points γ_1 and γ_2.

the convergence is not assured in general. Instead, we revert to a numerical method with global convergence to $\gamma_{\rm on}$. This is shown in Algorithm 3. Again, a convex optimization problem is solved at each iteration.

Algorithm 3

$\lambda_{min} = {\rm P}_{\rm D}^{'}(\gamma_1),\ \lambda_{max} = {\rm P}_{\rm D}^{'}(\gamma_2)$
$\gamma_{min} = \gamma_2,\ \gamma_{max} = \infty$
do
 $\lambda = (\lambda_{max} + \lambda_{min})/2$
 $\gamma_X = \underset{\gamma\in(\gamma_{min},\gamma_{max})}{\rm argmax} \quad {\rm P}_{\rm D}(\gamma) - \lambda\gamma$
 if ${\rm P}_{\rm D}(\gamma_X) - {\rm P}_{\rm D}^{'}(\gamma_X)\gamma_X > \alpha\,,$
 then $\lambda_{min} = \lambda,\ \gamma_{max} = \gamma_X$
 else $\lambda_{max} = \lambda,\ \gamma_{min} = \gamma_X$
while $\left|{\rm P}_{\rm D}(\gamma_X) - {\rm P}_{\rm D}^{'}(\gamma_X)\gamma_X - \alpha\right| > \epsilon$

Fig. 3 provides more insight about the near-optimal performance of the proposed approach. For various values of the false alarm probability α, we have computed the inflection points γ_1 and γ_2 from (6), evaluated the corresponding detection probabilities ${\rm P}_{\rm D}(\gamma_1)$ and ${\rm P}_{\rm D}(\gamma_2)$, respectively, and plotted the resulting detection performance curves with respect to α. As the false alarm constraint is tightened (smaller values), it is observed that the vertical gap between the detection performances calculated at the respective inflection points becomes much more pronounced. Since ${\rm P}_{\rm D}(\gamma)$ is monotonically increasing and $\gamma_{C1} \le \gamma_1$ is assured from Lemma 1, ${\rm P}_{\rm D}(\gamma_{C1})$ always takes values smaller than ${\rm P}_{\rm D}(\gamma_1)$, which is denoted with the red curve. On the contrary, the detection probability corresponding to the larger contact point γ_{C2} results in ${\rm P}_{\rm D}(\gamma_{C2}) \ge {\rm P}_{\rm D}(\gamma_2)$, which is represented by the blue curve. For a given α, the optimal strategy stated in Proposition 2 time shares between γ_{C1} and γ_{C2}, whose contributions to the detection performance should therefore lie below the red curve and above the blue curve, respectively. As a result, the contribution from the smaller contact point γ_{C1} can safely be ignored over a large set of false alarm probabilities without sacrificing from the detection performance claimed by the optimal strategy stated in Proposition 2. When the example in Section III.B is solved by assuming on-off signaling, it is observed that there is virtually no performance degradation.

D. Extension to Multidimensional Case

As mentioned earlier in the introduction, when the observations acquired by the receiver are corrupted with colored Gaussian noise, the detection probability can be maximized by transmitting along the eigenvector corresponding to the minimum eigenvalue of the noise covariance matrix [1]. More specifically, we consider the following hypothesis-testing problem where, given an M dimensional data vector, we have to decide between $\mathcal{H}_0 : \mathbf{Y} = \mathbf{N}$ and $\mathcal{H}_1 : \mathbf{Y} = \sqrt{S}\mathbf{v}_{min} + \mathbf{N}$, where $\mathbf{N} \sim \mathcal{N}(0, \mathbf{\Sigma})$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma}$, and $\mathbf{v}_{min}$ is the normalized eigenvector corresponding to the minimum eigenvalue of $\mathbf{\Sigma}$ with $|\mathbf{v}_{min}|^2 = 1$. It should be pointed out that a feedback mechanism is required from the receiver to the transmitter in order to facilitate signaling along the least noisy direction. In the absence of such a mechanism, the following analysis provides an upper bound on the detection performance.

At the receiver, the optimal correlation detector employs the decision statistics $T(\mathbf{y}) = \mathbf{v}_{min}^T \cdot \mathbf{y}$, which is a linear combination of jointly Gaussian random variables. Hence, the hypotheses can be rewritten as $\mathcal{H}_0 : T(\mathbf{Y}) \sim \mathcal{N}(0, \lambda_{min})$ and $\mathcal{H}_1 : T(\mathbf{Y}) \sim \mathcal{N}(\sqrt{S}, \lambda_{min})$, where λ_{min} denotes the minimum eigenvalue of $\mathbf{\Sigma}$ [1]. From the false alarm constraint, the detector threshold can be obtained as ${\rm P}_{\rm FA} = {\rm P}_0\left(T(\mathbf{Y}) \ge \eta\right) = Q\left(\eta/\sqrt{\lambda_{min}}\right) = \alpha$ and $\eta = \sqrt{\lambda_{min}}Q^{-1}(\alpha)$. The corresponding optimal NP decision rule is given as

$$\delta_{NP}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \mathbf{v}_{min}^T \cdot \mathbf{y} \ge \sqrt{\lambda_{min}}Q^{-1}(\alpha) \\ 0 & \text{if } \mathbf{v}_{min}^T \cdot \mathbf{y} < \sqrt{\lambda_{min}}Q^{-1}(\alpha) \end{cases} \qquad (10)$$

By defining $\gamma \triangleq S/\lambda_{min}$, the detection probability attained by δ_{NP} is computed from ${\rm P}_{\rm D}(\gamma) = {\rm P}_1(T(\mathbf{Y}) \ge \sqrt{\lambda_{min}}\,Q^{-1}(\alpha)) = Q\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)$. Notice that this expression is exactly in the same form as (4) after replacing σ^2 with λ_{min} and similar results to those in Section III can be obtained in this multidimensional setting.

IV. Convexity Properties in Noise Power

In this section, we investigate the binary hypothesis testing problem stated in (1) from the perspective of a power constrained jammer. By assuming signal power S to be fixed, we aim to determine the optimal power allocation strategy for a power constrained jammer that aims to minimize the detection probability at the receiver. The jamming noise is typically modeled with a Gaussian distribution [4], [8], [11], [12]. The power of the jammer is controlled over time through the variable σ^2, which is independent of S and N. It is assumed that the jamming power varies slowly in comparison with the sampling time at the receiver so that a smart receiver can estimate the current value of the jamming power σ^2 [12].[4] Then, the receiver updates its decision threshold via $\eta = \sigma Q^{-1}(\alpha)$ to maintain a constant false

[4]On the other hand, if the jamming power changes rapidly within the sampling period at the receiver, the net effect observed by the receiver would be jamming at the average power, which is shown to be suboptimal in Proposition 4 for jammers subject to stringent average power constraints.

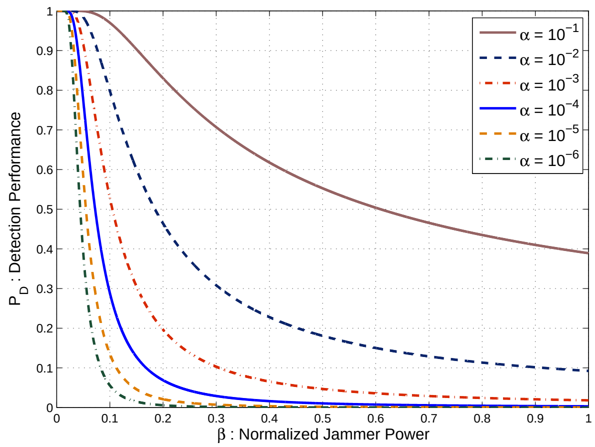

Fig. 4. Detection probability of the NP decision rule in (3) is plotted versus β for various values of the false alarm probability α. As an example, when $\alpha = 10^{-4}$, the inflection point is located at $\beta^* \approx 0.05164$ with $\mathrm{P_D}(\beta^*) \approx 0.7523$.

alarm probability α. Until the jammer changes its power to another value for σ^2, this is the optimal $\alpha-$level NP decision rule. On the other hand, jamming would be performed more effectively if the receiver could not adapt to varying jamming power.

Under constant transmit power S, the detection probability as a function of the normalized jamming power, $\beta \triangleq \sigma^2/S$, can be expressed as $\mathrm{P_D}(\beta) = Q\left(Q^{-1}(\alpha) - \beta^{-1/2}\right)$. The limits can be computed as $\lim_{\beta\to 0}\mathrm{P_D}(\beta) = 1$ and $\lim_{\beta\to\infty}\mathrm{P_D}(\beta) = \alpha$. Differentiating with respect to β yields $\mathrm{P}'_{\mathrm{D}}(\beta) = -(2\sqrt{2\pi})^{-1}\beta^{-3/2}\exp\Big\{-0.5\big(Q^{-1}(\alpha) - \beta^{-1/2}\big)^2\Big\}$, which is negative $\forall\,\beta > 0$. The limits for the first derivative are $\lim_{\beta\to 0}\mathrm{P}'_{\mathrm{D}}(\beta) = 0$ and $\lim_{\beta\to\infty}\mathrm{P}'_{\mathrm{D}}(\beta) = 0$.

Proposition 3: $\mathrm{P_D}(\beta)$ is a monotonically decreasing function of $\beta \in (0,\infty)$ with a single inflection point at

$$\beta^* = \left(\frac{\sqrt{(Q^{-1}(\alpha))^2 + 12} - Q^{-1}(\alpha)}{6}\right)^2 \tag{11}$$

that $\mathrm{P_D}(\beta)$ is strictly concave for $\beta < \beta^*$ and strictly convex for $\beta > \beta^*$.

Proof: The second derivative of the detection probability is $\mathrm{P}''_{\mathrm{D}}(\beta) = (4\sqrt{2\pi})^{-1}\beta^{-7/2}\exp\Big\{-0.5\big(Q^{-1}(\alpha) - \beta^{-1/2}\big)^2\Big\}\big(3\beta + Q^{-1}(\alpha)\sqrt{\beta} - 1\big)$. As before, the sign of the second derivative is determined by the right-most expression in parentheses. By substituting $x \triangleq \sqrt{\beta}$, the roots of the resulting quadratic polynomial are obtained as $\big(-Q^{-1}(\alpha) \pm \sqrt{(Q^{-1}(\alpha))^2 + 12}\big)/6$. Since $x = \sqrt{\beta} > 0$, the positive root results in the inflection point given in (11) indicating that $\mathrm{P_D}(\beta)$ is strictly concave for $\beta < \beta^*$ and strictly convex for $\beta > \beta^*$. □

The detection performance of the NP detector given by (3) is depicted in Fig. 4 versus β for various values of the false alarm probability α, which point out the possibility of decreasing the detection probability via time sharing of the jammer noise power. In order to obtain the optimal time sharing strategy for the jammer, we first present the following lemma which can be proved using a similar approach to that provided in the Appendix.

Lemma 3: Let β^* be the inflection point of $\mathrm{P_D}(\beta)$ as given in (11). There exists a unique point $\beta_C \geq \beta^*$ such that the tangent to $\mathrm{P_D}(\beta)$ at β_C lies below $\mathrm{P_D}(\beta)$ and passes through the point $(0,1)$.

The contact point β_C can be obtained from $\mathrm{P_D}(\beta_C) - \beta_C\mathrm{P}'_{\mathrm{D}}(\beta_C) = 1$, or equivalently solving for $\hat{x}$ in

$$\hat{x} = Q^{-1}\left(1 - \frac{Q^{-1}(\alpha) - \hat{x}}{2\sqrt{2\pi}}\exp\left\{-\frac{\hat{x}^2}{2}\right\}\right) \tag{12}$$

and then substituting into $\beta_C = (Q^{-1}(\alpha) - \hat{x})^{-2}$. A fixed point iteration approach is not guaranteed to converge in general. Fortunately, a variant of the proposed numerical method can be employed to obtain β_C as well. Once again, a convex optimization problem is solved at each iteration and the bisection search facilitates rapid convergence.

Algorithm 4

$\lambda_{min} = \mathrm{P}'_{\mathrm{D}}(\beta^*),\ \lambda_{max} = 0$
$\beta_{min} = \beta^*,\ \beta_{max} = \infty$
do
  $\lambda = (\lambda_{max} + \lambda_{min})/2$
  $\beta_X = \operatorname*{argmin}_{\beta\in(\beta_{min},\beta_{max})} \mathrm{P_D}(\beta) - \lambda\beta$
  if $\mathrm{P_D}(\beta_X) - \mathrm{P}'_{\mathrm{D}}(\beta_X)\beta_X > 1$,
  then $\lambda_{min} = \lambda,\ \beta_{min} = \beta_X$
  else $\lambda_{max} = \lambda,\ \beta_{max} = \beta_X$
while $\left|\mathrm{P_D}(\beta_X) - \mathrm{P}'_{\mathrm{D}}(\beta_X)\beta_X - 1\right| > \epsilon$

Next, we present the optimal strategy for a Gaussian jammer operating under peak power constraint $\mathrm{J_{peak}}$ and average power constraint $\mathrm{J_{avg}}$ ($\mathrm{J_{avg}} \leq \mathrm{J_{peak}}$) towards a smart receiver employing the adaptable threshold detector given in (3).

Proposition 4: The jammer's optimal strategy is to switch between powers 0 and β_C with the fraction of on-power time $\mathrm{J_{avg}}/\beta_C$ when $\mathrm{J_{avg}} < \beta_C < \mathrm{J_{peak}}$. For $\beta_C \geq \mathrm{J_{peak}}$, the optimal strategy time shares between powers 0 and $\mathrm{J_{peak}}$ with the fraction of on-power time $\mathrm{J_{avg}}/\mathrm{J_{peak}}$. For $\mathrm{J_{avg}} > \beta_C$, jamming is performed continuously at the average power.

Again the proof follows by noting that the stated strategy results in the largest convex function that is smaller than $\mathrm{P_D}(\beta)$ for $\beta \in [0, \mathrm{J_{peak}}]$. Finally as an example, for $\alpha = 10^{-4}$, $\mathrm{J_{avg}} = 0.04$, and $\mathrm{J_{peak}} = 0.1$, on-off Gaussian jamming can reduce the detection probability from 0.8999 down to 0.7109 by transmitting with power $\beta_C = 0.08779$ for approximately 45.56 percent of the time and aborting jamming for 54.44 percent of the time. If the peak power constraint is lowered to $\mathrm{J_{peak}} = 0.06$, the optimal strategy can still decrease the detection probability to 0.7612 by time sharing between 0 and peak power $\mathrm{J_{peak}} = 0.06$ with two-thirds of on-power time fraction.

V. Conclusions and Future Work

In this paper, we have examined the convexity properties of the detection probability for the problem of determining the presence of a target signal immersed in additive Gaussian noise. Unnoticed in the previous literature on the NP framework, we have found out that the detection performance of a power constrained transmitter can be increased via time sharing between different levels whenever the false alarm requirement is smaller than $Q(2) \approx 0.02275$. Although the optimal strategy indicates time sharing between two nonzero power levels for moderate values of the power constraint, it is shown that the on-off signaling strategy can well approximate the optimal performance. Next, we have considered the dual problem for a power constrained jammer and proved the existence of a critical power level up to which on-off jamming can be employed to degrade the detection performance of a smart receiver. A future work is to analyze how the optimal strategy for the transmitter changes with the jammer's time sharing and vice versa. Equilibrium conditions can be sought in a game-theoretic setting.

The results in this study can be applied for slow fading channels assuming that perfect channel state information (CSI) is present at the transmitter, and a short-term power constraint is imposed by computing the average over a time period close to the duration of the channel coherence time. In that case, the only modification in the formulations would be to update the definition of γ by scaling it with the channel power gain. In particular, considering a block fading channel model, the proposed optimal and suboptimal signaling approaches can be employed for each block. If the transmitter does not have perfect CSI, then the detection probability achieved by the optimal signaling approach based on perfect CSI can be used as an upper bound on the detection performance. For fast fading channels, the instantaneous CSI may not be available at the transmitter and the optimum power control strategy, which adapts the transmit power as a function of the instantaneous channel power gain, may not be obtained. The performance metric should be changed to the average detection probability over the fading distribution. In that case, the convexity properties would change (and in general depend on the fading distribution), and a new analysis would be required. Nevertheless, we can still state that the average detection probability is concave with respect to the transmit signal power for $\alpha \in [Q(2), 1)$ since a nonnegative weighted sum of concave functions is concave. Moreover, the optimal power control scheme can still be described as time sharing between *at most* two power levels due to Carathéodory's theorem [13], but whether the time sharing would improve over the constant power transmission scheme and over which regions it would improve need to be analyzed for the specific fading distribution under consideration.

APPENDIX

A. Proof of Lemma 1

As can be noted from the expression in the first paragraph of Section III.A, the derivative of the detection probability $\mathrm{P}_{\mathrm{D}}^{'}(\gamma)$ is a continuous and positive function $\forall \gamma > 0$ with the limits $\lim_{\gamma\to 0} \mathrm{P}_{\mathrm{D}}^{'}(\gamma) = \infty$ and $\lim_{\gamma\to\infty} \mathrm{P}_{\mathrm{D}}^{'}(\gamma) = 0$. In Proposition 1, it is stated that $\mathrm{P}_{\mathrm{D}}(\gamma)$ is strictly concave over the intervals $(0, \gamma_1)$ and (γ_2, ∞), whereas it is strictly convex over the interval (γ_1, γ_2). More precisely, $\mathrm{P}_{\mathrm{D}}^{'}(\gamma)$ monotonically decreases over the interval $(0, \gamma_1)$, monotonically increases over the interval (γ_1, γ_2), and monotonically decreases over the interval (γ_2, ∞). Therefore, there exists a unique point $\gamma_{1x} \in (0, \gamma_1]$, at which the derivative of the detection probability is equal to that at the second inflection point, i.e., $\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{1x}) = \mathrm{P}_{\mathrm{D}}^{'}(\gamma_2)$. Similarly, there exists a unique point $\gamma_{2x} \in [\gamma_2, \infty)$, at which the derivative of the detection probability is equal to that at the first inflection point, i.e., $\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{2x}) = \mathrm{P}_{\mathrm{D}}^{'}(\gamma_1)$. More generally, for every $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$ there exists a unique point $\hat{\gamma}_2 \in [\gamma_2, \gamma_{2x}]$ such that the derivatives at both points are equal $\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1) = \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_2)$. In other words, a one-to-one continuous function can be defined from the interval $[\gamma_{1x}, \gamma_1]$ onto the interval $[\gamma_2, \gamma_{2x}]$ as follows $\hat{\gamma}_2(\hat{\gamma}_1) = (\mathrm{P}_{\mathrm{D}}^{'})^{-1}\left(\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)\right)$. Now, consider the function $f(\gamma, \hat{\gamma}_1) \triangleq \mathrm{P}_{\mathrm{D}}(\gamma) - \left(\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)(\gamma - \hat{\gamma}_1) + \mathrm{P}_{\mathrm{D}}(\hat{\gamma}_1)\right)$, which provides the vertical difference between the detection probability $\mathrm{P}_{\mathrm{D}}(\gamma)$ and the value of the line tangent to the detection probability curve at $\hat{\gamma}_1$. Recall that for a given $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$, $\partial f/\partial\gamma = \mathrm{P}_{\mathrm{D}}^{'}(\gamma) - \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)$ is zero at a unique point $\hat{\gamma}_2 \in [\gamma_2, \gamma_{2x}]$. Next, we define the following continuous function: $h(\hat{\gamma}_1) \triangleq f(\hat{\gamma}_2(\hat{\gamma}_1), \hat{\gamma}_1) = \mathrm{P}_{\mathrm{D}}(\hat{\gamma}_2) - \mathrm{P}_{\mathrm{D}}(\hat{\gamma}_1) - \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1)$. The operation of this function can be described informally as follows. It takes as input a point $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$, finds the corresponding unique point $\hat{\gamma}_2 \in [\gamma_2, \gamma_{2x}]$ with the same slope such that $\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_2) = \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)$, draws the tangent line to the detection probability curve at the point $\hat{\gamma}_1$ with the slope $\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)$, and calculates the vertical separation between the detection probability curve and the tangent line at the point $\hat{\gamma}_2$. In the sequel, we show that $h(\cdot)$ has a unique root $\gamma_{C1} \in [\gamma_{1x}, \gamma_1]$. By differentiation, it is observed that $h(\cdot)$ is an increasing function over $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$. More formally, $\partial h(\hat{\gamma}_1)/\partial\hat{\gamma}_1 = \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_2)\hat{\gamma}_2' - \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1) - \mathrm{P}_{\mathrm{D}}^{''}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1) - \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1)(\hat{\gamma}_2' - 1) = -\mathrm{P}_{\mathrm{D}}^{''}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1) > 0$, where the last equality follows from $\mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_1) = \mathrm{P}_{\mathrm{D}}^{'}(\hat{\gamma}_2)$ and the inequality is due to the strict concavity of $\mathrm{P}_{\mathrm{D}}(\hat{\gamma}_1)$ over $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$. By selecting $\hat{\gamma}_1 = \gamma_{1x}$, we have $\hat{\gamma}_2 = \gamma_2$ and $h(\gamma_{1x}) = \mathrm{P}_{\mathrm{D}}(\gamma_2) - \mathrm{P}_{\mathrm{D}}(\gamma_{1x}) - \mathrm{P}_{\mathrm{D}}^{'}(\gamma_{1x})(\gamma_2 - \gamma_{1x}) \leq 0$. The last inequality follows by noting that $\mathrm{P}_{\mathrm{D}}^{'}(\gamma) \leq \mathrm{P}_{\mathrm{D}}^{'}(\gamma_{1x})$ for $\gamma \in [\gamma_{1x}, \gamma_2]$ and $\mathrm{P}_{\mathrm{D}}(\gamma_2) = \mathrm{P}_{\mathrm{D}}(\gamma_{1x}) + \int_{\gamma_{1x}}^{\gamma_2} \mathrm{P}_{\mathrm{D}}^{'}(\gamma)d\gamma$. On the other hand, by selecting $\hat{\gamma}_1 = \gamma_1$, we have $\hat{\gamma}_2 = \gamma_{2x}$ and $h(\gamma_1) = \mathrm{P}_{\mathrm{D}}(\gamma_{2x}) - \mathrm{P}_{\mathrm{D}}(\gamma_1) - \mathrm{P}_{\mathrm{D}}^{'}(\gamma_1)(\gamma_{2x} - \gamma_1) \geq 0$. Again, the inequality follows from $\mathrm{P}_{\mathrm{D}}^{'}(\gamma) \geq \mathrm{P}_{\mathrm{D}}^{'}(\gamma_1)$ for $\gamma \in [\gamma_1, \gamma_{2x}]$ and $\mathrm{P}_{\mathrm{D}}(\gamma_{2x}) = \mathrm{P}_{\mathrm{D}}(\gamma_1) + \int_{\gamma_1}^{\gamma_{2x}} \mathrm{P}_{\mathrm{D}}^{'}(\gamma)d\gamma$. Since $h(\cdot)$ is a continuous and increasing function, it must have a unique root $\gamma_{C1} \in [\gamma_{1x}, \gamma_1]$. Consequently, tangent to $\mathrm{P}_{\mathrm{D}}(\gamma)$ at γ_{C1} is also tangent at the point $\gamma_{C2} = (\mathrm{P}_{\mathrm{D}}^{'})^{-1}\left(\mathrm{P}_{\mathrm{D}}^{'}(\gamma_{C1})\right) \in [\gamma_2, \gamma_{2x}]$.

Next, we show that the tangent line, which passes through the points $(\gamma_{C1}, \mathrm{P}_{\mathrm{D}}(\gamma_{C1}))$ and $(\gamma_{C2}, \mathrm{P}_{\mathrm{D}}(\gamma_{C2}))$, lies above $\mathrm{P}_{\mathrm{D}}(\gamma)$ for all $\gamma > 0$. Since $\mathrm{P}_{\mathrm{D}}(\gamma)$ is strictly concave over $(0, \gamma_1)$, the tangent at γ_{C1} lies above $\mathrm{P}_{\mathrm{D}}(\gamma)$ for $\gamma \in (0, \gamma_1)$. Recall that the same line is also tangent to $\mathrm{P}_{\mathrm{D}}(\gamma)$ at γ_{C2} and as a result, it lies above $\mathrm{P}_{\mathrm{D}}(\gamma)$ for $\gamma > \gamma_2$. Subsequently, the line segment connecting the points $(\gamma_1, \mathrm{P}_{\mathrm{D}}(\gamma_1))$ and $(\gamma_2, \mathrm{P}_{\mathrm{D}}(\gamma_2))$ lies above $\mathrm{P}_{\mathrm{D}}(\gamma)$ for $\gamma \in [\gamma_1, \gamma_2]$ since $\mathrm{P}_{\mathrm{D}}(\gamma)$ is convex over this interval. Since the inflection points $(\gamma_1, \mathrm{P}_{\mathrm{D}}(\gamma_1))$ and $(\gamma_2, \mathrm{P}_{\mathrm{D}}(\gamma_2))$ are below the tangent line, the line segment connecting them also lies below the tangent line. This proves that the tangent line lies above $\mathrm{P}_{\mathrm{D}}(\gamma)$ for all $\gamma > 0$. □

REFERENCES

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[2] R. G. Gallager, *Principles of Digital Communication*. Cambridge, UK: Cambridge University Press, 2008.

[3] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Trans. Info. Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.

[4] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. Info. Theory*, vol. 56, pp. 1948–1960, Apr. 2010.

[5] C. Goken, S. Gezici, and O. Arikan, “Optimal stochastic signaling for power-constrained binary communications systems,” *IEEE Trans. Wirel. Commun.*, vol. 9, no. 12, pp. 3650 –3661, Dec. 2010.

[6] B. Dulek and S. Gezici, “Detector randomization and stochastic signaling for minimum probability of error receivers,” *IEEE Tran. Commun.*, vol. 60, no. 4, pp. 923–928, April 2012.

[7] ——, “Optimal signaling and detector design for power constrained on-off keying systems in Neyman-Pearson framework,” in *IEEE 16th Wksp. Stat. Signal Process.*, June 2011, pp. 93–96.

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

[9] A. Mukherjee and A. L. Swindlehurst, “Prescient beamforming in multi-user interweave cognitive radio networks,” in *4th IEEE Int. Wksp. Comput. Adv. Multi-Sensor Adap. Process. (CAMSAP’11)*, Dec. 2011, pp. 253–256.

[10] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Singapore: Athena Scientific, 1995.

[11] W. E. S. R. J. McEliece, “An information theoretic study of communication in the presence of jamming,” in *Int. Conf. Commun. (ICC’81)*, vol. 3, 1981, p. 45.

[12] M. Weiss and S. C. Schwartz, “On optimal minimax jamming and detection of radar signals,” *IEEE Trans. Aeros. Elect. Sys.*, vol. AES-21, no. 3, pp. 385–393, May 1985.

[13] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1968.