

Optimal Stochastic Signaling for Power-Constrained Binary Communications Systems

Cagri Goken, *Student Member, IEEE*, Sinan Gezici, *Member, IEEE*, and Orhan Arıkan, *Member, IEEE*

Abstract—Optimal stochastic signaling is studied under second and fourth moment constraints for the detection of scalar-valued binary signals in additive noise channels. Sufficient conditions are obtained to specify when the use of stochastic signals instead of deterministic ones can or cannot improve the error performance of a given binary communications system. Also, statistical characterization of optimal signals is presented, and it is shown that an optimal stochastic signal can be represented by a randomization of at most three different signal levels. In addition, the power constraints achieved by optimal stochastic signals are specified under various conditions. Furthermore, two approaches for solving the optimal stochastic signaling problem are proposed; one based on particle swarm optimization (PSO) and the other based on convex relaxation of the original optimization problem. Finally, simulations are performed to investigate the theoretical results, and extensions of the results to M -ary communications systems and to other criteria than the average probability of error are discussed.

Index Terms—Detection, binary communications, additive noise channels, randomization, probability of error, optimization.

I. INTRODUCTION

In this paper, optimal signaling techniques are investigated for minimizing the average probability of error of a binary communications system under power constraints. Optimal signaling in the presence of zero-mean Gaussian noise has been studied extensively in the literature [1], [2]. It is shown that deterministic antipodal signals, i.e., $S_1 = -S_0$, minimize the average probability of error of a binary communications system in additive Gaussian noise channels when the average power of each signal is constrained by the same limit. In addition, for vector observations, selecting the deterministic signals along the eigenvector of the covariance matrix of the Gaussian noise corresponding to the minimum eigenvalue minimizes the average probability of error under power constraints in the form of $\|S_0\|^2 \leq A$ and $\|S_1\|^2 \leq A$ [2, pp. 61–63]. In [3], optimal binary communications over additive white Gaussian noise (AWGN) channels are studied for nonequal prior probabilities under an average energy per bit constraint. It is shown that the optimal signaling scheme is on-off keying (OOK) for coherent detection when the signals have nonnegative correlation, and that the optimal signaling is OOK also for envelope detection for any signal correlation. In [4], a source-controlled turbo coding algorithm is proposed for nonuniform binary memoryless sources over AWGN channels by utilizing asymmetric nonbinary signal constellations. Although the average probability of error expressions and

optimal signaling techniques are well-known when the noise is Gaussian, the noise can have significantly different probability distribution than the Gaussian distribution in some cases due to effects such as multiuser interference and jamming [5]–[7]. In [8], additive noise channels with binary inputs and scalar outputs are studied, and the worst-case noise distribution is characterized. Specifically, it is shown that the least-favorable noise distribution that maximizes the average probability of error and minimizes the channel capacity is a mixture of discrete lattices [8]. A similar problem is considered in [9] for a binary communications system in the presence of an additive jammer, and properties of optimal jammer distribution and signal distribution are obtained.

In [6], the convexity properties of the average probability of error are investigated for binary-valued scalar signals over additive noise channels under an average power constraint. It is shown that the average probability of error is a convex nonincreasing function for unimodal differentiable noise probability density functions (PDFs) when the receiver employs maximum likelihood (ML) detection. Based on this result, it is concluded that randomization of signal values (or, stochastic signal design) cannot improve error performance for the considered communications system. Then, the problem of maximizing the average probability of error is studied for an average power-constrained jammer, and it is shown that the optimal solution can be obtained when the jammer randomizes its power between at most two power levels. Finally, the results are applied to multiple additive noise channels, and optimum channel switching strategy is obtained as time-sharing between at most two channels and power levels [6].

Optimal randomization between two deterministic signal pairs and the corresponding ML decision rules is studied in [10] for an average power-constrained antipodal binary communications system, and it is shown that power randomization can result in significant performance improvement. In [11], the problem of pricing and transmission scheduling is investigated for an access point in a wireless network, and it is proven that the randomization between two business decision and price pairs maximizes the time-average profit of the access point. Although the problem studied in [11] is in a different context, its theoretical approach is similar to those in [6] and [10] for obtaining optimal signal distributions.

Although the average probability of error of a binary communications system is minimized by *deterministic* antipodal signals in additive Gaussian noise channels [2], the studies in [6], [9], [10], [11] imply that *stochastic* signaling can sometimes achieve lower average probability of error when the noise is non-Gaussian. Therefore, a more generic formulation of the optimal signaling problem for binary communications systems can be stated as obtaining the optimal probability distributions of signals S_0 and S_1 such that the average

The authors 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: {goken,gezici,arikan}@ee.bilkent.edu.tr. Part of this research was presented at the IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), Marrakech, Morocco, June 2010.

probability of error of the system is minimized under certain constraints on the moments of S_0 and S_1 . It should be noted that the main difference of this optimal stochastic signaling approach from the conventional (deterministic) approach [1], [2] is that signals S_0 and S_1 are considered as random variables in the former whereas they are regarded as deterministic quantities in the latter.

Although randomization between deterministic signal constellations and corresponding optimal detectors is studied in an additive Gaussian mixture noise channel under an average power constraint in [10], no studies have considered the optimal stochastic signaling problem based on a generic formulation (i.e., for arbitrary receivers and noise probability distributions) under *both* average power *and* peakedness constraints on individual signals. In this paper, such a generic formulation of the stochastic signaling problem is considered, and sufficient conditions for improvability and nonimprovability of error performance via stochastic signal design are derived. In addition, the statistical characterization of optimal signals is provided and two optimization theoretic approaches are proposed for obtaining the optimal signals. The main contributions of the paper can be summarized as follows:

- Formulation of the optimal stochastic signaling problem under both average power and peakedness constraints.
- Derivation of sufficient conditions to determine whether stochastic signaling can provide error performance improvement compared to the conventional (deterministic) signaling.
- Statistical characterization of optimal signals, which reveals that an optimal stochastic signal can be expressed as a randomization of at most three different signals levels.
- Study of two optimization techniques, namely particle swarm optimization (PSO) [12] and convex relaxation [13], in order to obtain optimal and close-to-optimal solutions to the stochastic signaling problem.

In addition to the results listed above, the power constraints achieved by optimal signals are specified under various conditions. Also, simulation results are presented to investigate the theoretical results. Finally, it is explained that the results obtained for minimizing the average probability of error for a binary communications system can be extended to M -ary systems, as well as to other performance criteria than the average probability of error, such as the Bayes risk [2], [14].

II. SYSTEM MODEL AND MOTIVATION

Consider a scalar binary communications system, as in [6], [8] and [15], in which the received signal is expressed as

$$Y = S_i + N, \quad i \in \{0, 1\}, \quad (1)$$

where S_0 and S_1 represent the transmitted signal values for symbol 0 and symbol 1, respectively, and N is the noise component that is independent of S_i . In addition, the prior probabilities of the symbols, which are represented by π_0 and π_1 , are assumed to be known.

As stated in [6], the scalar channel model in (1) provides an abstraction for a continuous-time system that processes the received signal by a linear filter and samples it once per symbol interval. In addition, although the signal model in (1) is in the form of a simple additive noise channel, it also holds for flat-fading channels assuming perfect channel estimation.

In that case, the signal model in (1) can be obtained after appropriate equalization [1].

It should be noted that the probability distribution of the noise component in (1) is not necessarily Gaussian. Due to interference, such as multiple-access interference, the noise component can have a significantly different probability distribution from the Gaussian distribution [5], [6], [16].

A generic decision rule is considered at the receiver to determine the symbol in (1). That is, for a given observation $Y = y$, the decision rule $\phi(y)$ is specified as

$$\phi(y) = \begin{cases} 0, & y \in \Gamma_0 \\ 1, & y \in \Gamma_1 \end{cases}, \quad (2)$$

where Γ_0 and Γ_1 are the decision regions for symbol 0 and symbol 1, respectively [2].

The aim is to design signals S_0 and S_1 in (1) in order to minimize the average probability of error for a given decision rule, which is expressed as

$$P_{\text{avg}} = \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0), \quad (3)$$

where $P_i(\Gamma_j)$ is the probability of selecting symbol j when symbol i is transmitted. In practical systems, there are constraints on the average power and the peakedness of signals, which can be expressed as [17]

$$E\{|S_i|^2\} \leq A, \quad E\{|S_i|^4\} \leq \kappa A^2, \quad (4)$$

for $i = 0, 1$, where A is the average power limit and the second constraint imposes a limit on the peakedness of the signal depending on the $\kappa \in (1, \infty)$ parameter.¹ Therefore, the average probability of error in (3) needs to be minimized under the second and fourth moment constraints in (4).

The main motivation for the optimal stochastic signaling problem is to improve the error performance of the communications system by considering the signals at the transmitter as random variables and finding the optimal probability distributions for those signals [6]. Therefore, the generic problem can be formulated as obtaining the optimal probability distributions of the signals S_0 and S_1 for a given decision rule at the receiver under the average power and peakedness constraints in (4).

Since the optimal signal design is performed at the transmitter, the transmitter is assumed to have the knowledge of the statistics of the noise at the receiver and the channel state information. Although this assumption may not hold in some cases, there are certain scenarios in which it can be realized.² Consider, for example, the downlink of a multiple-access communications system, in which the received signal can be modeled as $Y = S^{(1)} + \sum_{k=2}^K \xi_k S^{(k)} + \eta$, where $S^{(k)}$ is the signal of the k th user, ξ_k is the correlation coefficient between user 1 and user k , and η is a zero-mean Gaussian noise component. For the desired signal component $S^{(1)}$, $N = \sum_{k=2}^K \xi_k S^{(k)} + \eta$ forms the total noise, which has Gaussian mixture distribution. When the receiver sends via feedback the variance of noise η and the signal-to-noise ratio (SNR) to the transmitter, the transmitter can fully characterize

¹Note that for $E\{|S_i|^2\} = A$, the second constraint becomes $E\{|S_i|^4\}/(E\{|S_i|^2\})^2 \leq \kappa$, which limits the kurtosis of the signal [17].

²As discussed in Section VI, the problem studied in this paper can be considered for other systems than communications; hence, the practicality of the assumption depends on the specific application domain.

the PDF of the total noise N , as it knows the transmitted signal levels of all the users and the correlation coefficients.

In the conventional signal design, S_0 and S_1 are considered as deterministic signals, and they are set to $S_0 = -\sqrt{A}$ and $S_1 = \sqrt{A}$ [1], [2]. In that case, the average probability of error expression in (3) becomes

$$P_{\text{avg}}^{\text{conv}} = \pi_0 \int_{\Gamma_1} p_N(y + \sqrt{A}) dy + \pi_1 \int_{\Gamma_0} p_N(y - \sqrt{A}) dy, \quad (5)$$

where $p_N(\cdot)$ is the PDF of the noise in (1). As investigated in Section III-A, the conventional signal design is optimal for certain classes of noise PDFs and decision rules. However, in some cases, use of stochastic signals instead of deterministic ones can improve the system performance. In the following section, conditions for optimality and suboptimality of the conventional signal design are derived, and properties of optimal signals are investigated.

III. OPTIMAL STOCHASTIC SIGNALING

Instead of employing constant levels for S_0 and S_1 as in the conventional case, consider a more generic scenario in which the signal components can be stochastic. The aim is to obtain the optimal PDFs for S_0 and S_1 in (1) that minimize the average probability of error under the constraints in (4).

Let $p_{S_0}(\cdot)$ and $p_{S_1}(\cdot)$ represent the PDFs for S_0 and S_1 , respectively. Then, the average probability of error for the decision rule in (2) can be expressed from (3) as

$$P_{\text{avg}}^{\text{stoc}} = \pi_0 \int_{-\infty}^{\infty} p_{S_0}(t) \int_{\Gamma_1} p_N(y - t) dy dt + \pi_1 \int_{-\infty}^{\infty} p_{S_1}(t) \int_{\Gamma_0} p_N(y - t) dy dt. \quad (6)$$

Therefore, the optimal stochastic signal design problem can be stated as

$$\begin{aligned} & \min_{p_{S_0}, p_{S_1}} P_{\text{avg}}^{\text{stoc}} \\ & \text{subject to } E\{|S_i|^2\} \leq A, \quad E\{|S_i|^4\} \leq \kappa A^2, \quad i = 0, 1. \end{aligned} \quad (7)$$

Note that there are also implicit constraints in the optimization problem in (7), since $p_{S_i}(t)$ represents a PDF. Namely, $p_{S_i}(t) \geq 0 \forall t$ and $\int_{-\infty}^{\infty} p_{S_i}(t) dt = 1$ should also be satisfied by the optimal solution.

Since the aim is to obtain optimal stochastic signals for a given receiver, the decision rule in (2) is fixed (i.e., predefined Γ_0 and Γ_1). Therefore, the structure of the objective function $P_{\text{avg}}^{\text{stoc}}$ in (6) and the individual constraints on each signal imply that the optimization problem in (7) can be expressed as two decoupled optimization problems (see Appendix A). For example, the optimal signal for symbol 1 can be obtained from the solution of the following optimization problem:

$$\begin{aligned} & \min_{p_{S_1}} \int_{-\infty}^{\infty} p_{S_1}(t) \int_{\Gamma_0} p_N(y - t) dy dt \\ & \text{subject to } E\{|S_1|^2\} \leq A, \quad E\{|S_1|^4\} \leq \kappa A^2. \end{aligned} \quad (8)$$

A similar problem can be formulated for S_0 as well. Since the signals can be designed separately, the remainder of the paper focuses on the design of optimal S_1 according to (8).

The objective function in (8) can be expressed as the expectation of

$$G(S_1) \triangleq \int_{\Gamma_0} p_N(y - S_1) dy \quad (9)$$

over the PDF of S_1 . Then, the optimization problem in (8) becomes

$$\begin{aligned} & \min_{p_{S_1}} E\{G(S_1)\} \\ & \text{subject to } E\{|S_1|^2\} \leq A, \quad E\{|S_1|^4\} \leq \kappa A^2. \end{aligned} \quad (10)$$

It is noted that (10) provides a generic formulation that is valid for any noise PDF and detector structure. In the following sections, the signal subscripts are dropped for notational simplicity. Note that $G(x)$ in (9) represents the probability of deciding symbol 0 instead of symbol 1 when signal S_1 takes a constant value of x ; that is, $S_1 = x$.

A. On the Optimality of the Conventional Signaling

Under certain circumstances, using the conventional signaling approach, i.e., setting $S = \sqrt{A}$ (or, $p_S(x) = \delta(x - \sqrt{A})$), solves the optimization problem in (10). For example, if $G(x)$ achieves its minimum at $x = \sqrt{A}$; that is, $\arg \min_x G(x) = \sqrt{A}$, then $p_S(x) = \delta(x - \sqrt{A})$ becomes the optimal solution since it yields the minimum value for $E\{G(S_1)\}$ and also satisfies the constraints. However, this case is not very common as $G(x)$, which is the probability of deciding symbol 0 instead of symbol 1 when $S = x$, is usually a decreasing function of x ; that is, when a larger signal value x is used, smaller error probability can be obtained. Therefore, the following more generic condition is derived for the optimality of the conventional algorithm.

Proposition 1: *If $G(x)$ is a strictly convex and monotone decreasing function, then $p_S(x) = \delta(x - \sqrt{A})$ solves the optimization problem in (10).*

Proof: The proof is obtained via contradiction. First, it is assumed that there exists a PDF $p_{S_2}(x)$ for signal S that makes the conventional solution suboptimal; that is, $E\{G(S)\} < G(\sqrt{A})$ under the constraints in (10).

Since $G(x)$ is a strictly convex function, Jensen's inequality implies that $E\{G(S)\} > G(E\{S\})$. Therefore, as $G(x)$ is a monotone decreasing function, $E\{S\} > \sqrt{A}$ must be satisfied in order for $E\{G(S)\} < G(\sqrt{A})$ to hold true.

On the other hand, Jensen's inequality also states that $E\{S\} > \sqrt{A}$ implies $E\{S^2\} > (E\{S\})^2 > A$; that is, the constraint on the average power is violated (see (10)). Therefore, it is proven that no PDF can provide $E\{G(S)\} < G(\sqrt{A})$ and satisfy the constraints under the assumptions in the proposition. \square

As an example application of Proposition 1, consider a zero-mean Gaussian noise N in (1) with $p_N(x) = \exp(-x^2/(2\sigma^2))/\sqrt{2\pi}\sigma$, and a decision rule of the form $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$; i.e., the sign detector. Then, $G(x)$ in (9) can be obtained as

$$G(x) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right) dy = Q\left(\frac{x}{\sigma}\right), \quad (11)$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^{\infty} \exp(-t^2/2) dt$ defines the Q -function. It is observed that $G(x)$ in (11) is a monotone

decreasing and strictly convex function for $x > 0$.³ Therefore, the optimal signal is specified by $p_S(x) = \delta(x - \sqrt{A})$ from Proposition 1. Similarly, the optimal signal for symbol 0 can be obtained as $p_S(x) = \delta(x + \sqrt{A})$. Hence, the conventional signaling is optimal in this scenario.

B. Sufficient Conditions for Improvability

In this section, the aim is to determine when it is possible to improve the performance of the conventional signaling approach via stochastic signaling. A simple observation of (10) reveals that if the minimum of $G(x) = \int_{\Gamma_0} p_N(y - x) dy$ is achieved at x_{\min} with $x_{\min}^2 < A$, then $p_S(x) = \delta(x - x_{\min})$ becomes a better solution than the conventional one. In other words, if the noise PDF is such that the probability of selecting symbol 0 instead of symbol 1 is minimized for a signal value of $S_1 = x_{\min}$ with $x_{\min}^2 < A$, then the conventional solution can be improved. Another sufficient condition for the conventional algorithm to be suboptimal is to have a positive first-order derivative of $G(x)$ at $x = \sqrt{A}$, which can also be expressed from (9) as $-\int_{\Gamma_0} p'_N(y - \sqrt{A}) dy > 0$, where $p'_N(\cdot)$ denotes the derivative of $p_N(\cdot)$. In this case, $p_{S_2}(x) = \delta(x - \sqrt{A} + \epsilon)$ yields a smaller average probability of error than the conventional solution for infinitesimally small $\epsilon > 0$ values.

Although both of the conditions above are sufficient for improvability of the conventional algorithm, they are rarely met in practice since $G(x)$ is commonly a decreasing function of x as discussed before. Therefore, in the following, a sufficient condition is derived for more generic and practical conditions.

Proposition 2: *Assume that $G(x)$ is twice continuously differentiable around $x = \sqrt{A}$. Then, if $\int_{\Gamma_0} (p''_N(y - \sqrt{A}) + p'_N(y - \sqrt{A})/\sqrt{A}) dy < 0$ is satisfied, $p_S(x) = \delta(x - \sqrt{A})$ is not an optimal solution to (10).*

Proof: It is first observed from (9) that the condition in the proposition is equivalent to $G''(\sqrt{A}) < G'(\sqrt{A})/\sqrt{A}$. Therefore, in order to prove the suboptimality of the conventional solution $p_S(x) = \delta(x - \sqrt{A})$, it is shown that when $G''(\sqrt{A}) < G'(\sqrt{A})/\sqrt{A}$, there exists $\lambda \in (0, 1)$, $\epsilon > 0$ and $\Delta > 0$ such that $p_{S_2}(x) = \lambda \delta(x - \sqrt{A} + \epsilon) + (1 - \lambda) \delta(x - \sqrt{A} - \Delta)$ has a lower error probability than $p_S(x)$ while satisfying all the constraints in (10). More specifically, the existence of $\lambda \in (0, 1)$, $\epsilon > 0$ and $\Delta > 0$ that satisfy

$$\lambda G(\sqrt{A} - \epsilon) + (1 - \lambda) G(\sqrt{A} + \Delta) < G(\sqrt{A}) \quad (12)$$

$$\lambda(\sqrt{A} - \epsilon)^2 + (1 - \lambda)(\sqrt{A} + \Delta)^2 = A \quad (13)$$

$$\lambda(\sqrt{A} - \epsilon)^4 + (1 - \lambda)(\sqrt{A} + \Delta)^4 \leq \kappa A^2 \quad (14)$$

is sufficient to prove the suboptimality of the conventional signal design.

From (13), the following equation is obtained.

$$\lambda \epsilon^2 + (1 - \lambda) \Delta^2 = -2\sqrt{A} [(1 - \lambda) \Delta - \lambda \epsilon] . \quad (15)$$

³It is sufficient to consider the positive signal values only since $G(x)$ is monotone decreasing and the constraints x^2 and x^4 are even functions. In other words, no negative signal value can be optimal since its absolute value has the same constraint value but smaller $G(x)$.

If infinitesimally small ϵ and Δ values are selected, (12) can be approximated as

$$\begin{aligned} \lambda \left[G(\sqrt{A}) - \epsilon G'(\sqrt{A}) + \frac{\epsilon^2}{2} G''(\sqrt{A}) \right] + (1 - \lambda) \left[G(\sqrt{A}) \right. \\ \left. + \Delta G'(\sqrt{A}) + \frac{\Delta^2}{2} G''(\sqrt{A}) \right] < G(\sqrt{A}) \\ G'(\sqrt{A}) [(1 - \lambda) \Delta - \lambda \epsilon] + \frac{G''(\sqrt{A})}{2} [\lambda \epsilon^2 + (1 - \lambda) \Delta^2] < 0 \end{aligned} \quad (16)$$

When the condition in (15) is employed, (16) becomes

$$[(1 - \lambda) \Delta - \lambda \epsilon] \left(G'(\sqrt{A}) - \sqrt{A} G''(\sqrt{A}) \right) < 0 . \quad (17)$$

Since $(1 - \lambda) \Delta - \lambda \epsilon$ is always negative as can be noted from (15), the $G'(\sqrt{A}) - \sqrt{A} G''(\sqrt{A})$ term in (17) must be positive to satisfy the condition. In other words, when $G''(\sqrt{A}) < G'(\sqrt{A})/\sqrt{A}$, $p_{S_2}(x)$ can have a smaller error value than that of the conventional algorithm for infinitesimally small ϵ and Δ values that satisfy (15). To complete the proof, the condition in (14) needs to be verified for the specified ϵ and Δ values. From (15), (14) can be expressed, after some manipulation, as

$$\begin{aligned} A^2 + 16A\sqrt{A} [(1 - \lambda) \Delta - \lambda \epsilon] - 4\sqrt{A} [\lambda \epsilon^3 - (1 - \lambda) \Delta^3] \\ + [\lambda \epsilon^4 - (1 - \lambda) \Delta^4] \leq \kappa A^2 . \end{aligned} \quad (18)$$

Since $(1 - \lambda) \Delta - \lambda \epsilon$ is negative, the inequality can be satisfied for infinitesimally small ϵ and Δ , for which the third and the fourth terms on the left-hand-side become negligible compared to the first two. \square

The condition in Proposition 2 can be expressed more explicitly in practice. For example, if Γ_0 is the form of an interval, say $[\tau_1, \tau_2]$, then the condition in the proposition becomes $p_{N'}(\tau_2 - \sqrt{A}) - p_{N'}(\tau_1 - \sqrt{A}) + (p_N(\tau_2 - \sqrt{A}) - p_N(\tau_1 - \sqrt{A}))/\sqrt{A} < 0$. This inequality can be generalized in a straightforward manner when Γ_0 is the union of multiple intervals.

Since the condition in Proposition 2 is equivalent to $G''(\sqrt{A}) < G'(\sqrt{A})/\sqrt{A}$ (see (9)), the intuition behind the proposition can be explained as follows. As the optimization problem in (10) aims to minimize $E\{G(S)\}$ while keeping $E\{S^2\}$ and $E\{S^4\}$ below thresholds A and κA^2 , respectively, a better solution than $p_S(x) = \delta(x - \sqrt{A})$ can be obtained with multiple mass points if $G(x)$ is decreasing at an increasing rate (i.e., with a negative second derivative) such that an increase from $x = \sqrt{A}$ causes a fast decrease in $G(x)$ but relatively slow increase in x^2 and x^4 , and a decrease from $x = \sqrt{A}$ causes a fast decrease in x^2 and x^4 but relatively slow increase in $G(x)$. In that case, it becomes possible to use a PDF with multiple mass points and to obtain a smaller $E\{G(S)\}$ while satisfying $E\{S^2\} \leq A$ and $E\{S^4\} \leq \kappa A^2$.

Proposition 2 provides a simple sufficient condition to determine if there is any possibility for performance improvement over the conventional signal design. For a given noise PDF and a decision rule, the condition in Proposition 2 can be evaluated in a straightforward manner. In order to provide an illustrative example, consider the noise PDF

$$p_N(y) = \begin{cases} y^2, & |y| \leq 1.1447 \\ 0, & |y| > 1.1447 \end{cases} , \quad (19)$$

and a sign detector at the receiver; that is, $\Gamma_0 = (-\infty, 0]$. Then, the condition in Proposition 2 can be evaluated as

$$p_N'(-\sqrt{A}) + p_N(-\sqrt{A})/\sqrt{A} < 0. \quad (20)$$

Assuming that the average power is constrained to $A = 0.64$, the inequality in (20) becomes $2(-0.8) + (-0.8)^2/0.8 < 0$. Hence, Proposition 2 implies that the conventional solution is not optimal for this problem. For example, $p_S(x) = 0.391 \delta(x-0.988) + 0.333 \delta(x-0.00652) + 0.276 \delta(x-0.9676)$ yields an average error probability of 0.2909 compared to 0.3293 corresponding to the conventional solution $p_S(x) = \delta(x-0.8)$, as studied in Section IV.

Although the noise PDF in (19) is not common in practice, improvements over the conventional algorithm are possible and Proposition 2 can be applied also for certain types of Gaussian mixture noise (see Section IV), which is observed more frequently in practical scenarios [16]-[19]. For example, in multiuser wireless communications, the desired signal is corrupted by interfering signals from other users as well as zero-mean Gaussian noise, which altogether result in Gaussian mixture noise [16].

C. Statistical Characteristics of Optimal Signals

In this section, PDFs of optimal signals are characterized and it is shown that an optimal signal can be represented by a randomization of at most three different signal levels. In addition, it is proven that the optimal signal achieves at least one of the second and fourth moment constraints in (10) for most practical cases.

In the following proposition, it is stated that, in most practical scenarios, an optimal stochastic signal can be represented by a discrete random variable with no more than three mass points.

Proposition 3: Assume that the possible signal values are specified by $|S| \leq \gamma$ for a finite $\gamma > 0$, and $G(\cdot)$ in (9) is continuous. Then, an optimal solution to (10) can be expressed in the form of $p_S(x) = \sum_{i=1}^3 \lambda_i \delta(x-x_i)$, where $\sum_{i=1}^3 \lambda_i = 1$ and $\lambda_i \geq 0$ for $i = 1, 2, 3$.

Proof: Please see Appendix B.

The assumption in the proposition, which states that the possible signal values belong to set $[-\gamma, \gamma]$, is realistic for practical communications systems since arbitrarily large positive and negative signal values cannot be generated at the transmitter. In addition, for most practical scenarios, $G(\cdot)$ in (9) is continuous since the noise at the receiver, which is commonly the sum of zero-mean Gaussian thermal noise and interference terms that are independent from the thermal noise, has a continuous PDF.

The result in Proposition 3 can be extended to the problems with more constraints. Let $E\{G(S)\}$ be the objective function to minimize over possible PDFs $p_S(x)$, subject to $E\{H_i(S)\} \leq A_i$ for $i = 1, \dots, N_c$. Then, under the conditions in the proposition, the proof in Appendix B implies that there exists an optimal PDF with at most $N_c + 1$ mass points.⁴

The significance of Proposition 3 lies in the fact that it reduces the optimization problem in (10) from the space of all PDFs that satisfy the second and fourth moment constraints to the space of discrete PDFs with at most 3 mass points that

satisfy the second and fourth moment constraints. In other words, instead of optimization over functions, an optimization over a vector of 6 elements (namely, 3 mass point locations and their weights) can be considered for the optimal signaling problem as a result of Proposition 3. In addition, this result facilitates a convex relaxation of the optimization problem in (10) for any noise PDF and decision rule as studied in Section III-D.

Next, the second and the fourth moments of the optimal signals are investigated. Let x_{\min} represent the signal level that yields the minimum value of $G(x)$ in (9); that is, $x_{\min} = \arg \min_x G(x)$. If $x_{\min} < \sqrt{A}$, the optimal signal has the constant value of x_{\min} and the second and fourth moments are given by $x_{\min}^2 < A$ and $x_{\min}^4 < \kappa A^2$, respectively. However, it is more common to have $x_{\min} > \sqrt{A}$ since larger signal values are expected to reduce $G(x)$ as discussed before. In that case, the following proposition states that at least one of the constraints in (10) is satisfied.

Proposition 4: Let $x_{\min} = \arg \min_x G(x)$ be the unique minimum of $G(x)$.

a) If $A^2 < x_{\min}^4 < \kappa A^2$, then the optimal signal satisfies $E\{S^2\} = A$.

b) If $x_{\min}^4 > \kappa A^2$, then the optimal signal satisfies at least one of $E\{S^2\} = A$ and $E\{S^4\} = \kappa A^2$.

Proof: Please see Appendix C.

An important implication of Proposition 4 is that when $x_{\min} > \sqrt{A}$, any solution that results in second and fourth moments that are smaller than A and κA^2 , respectively, cannot be optimal. In other words, it is possible to improve that solution by increasing the second and/or the fourth moment of the signal until at least one of the constraints become active.

After characterizing the structure and the properties of optimal signals, two approaches are proposed in the next section to obtain optimal and close-to-optimal signal PDFs.

D. Calculation of the Optimal Signal

In order to obtain the PDF of an optimal signal, the constrained optimization problem in (10) should be solved. In this section, two approaches are studied in order to obtain optimal and close-to-optimal solutions to that optimization problem.

1) *Global Optimization Approach:* Since Proposition 3 states that the optimal signaling problem in (10) can be solved over PDFs in the form of $p_S(x) = \sum_{j=1}^3 \lambda_j \delta(x-x_j)$, (10) can be expressed as

$$\begin{aligned} \min_{\lambda, \mathbf{x}} \quad & \sum_{j=1}^3 \lambda_j G(x_j) \\ \text{subject to} \quad & \sum_{j=1}^3 \lambda_j x_j^2 \leq A, \quad \sum_{j=1}^3 \lambda_j x_j^4 \leq \kappa A^2, \\ & \sum_{j=1}^3 \lambda_j = 1, \quad \lambda_j \geq 0 \quad \forall j, \end{aligned} \quad (21)$$

where $\mathbf{x} = [x_1 \ x_2 \ x_3]^T$ and $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \lambda_3]^T$.

Note that the optimization problem in (21) is a not convex problem in general due to both the objective function and the first two constraints. Therefore, global optimization techniques, such as PSO, differential evolution and genetic

⁴It is assumed that $H_1(x), \dots, H_{N_c}(x)$ are bounded functions for the possible values of the signal.

algorithms [20] should be employed to obtain the optimal PDF. In this paper, the PSO approach [12], [21]-[23] is used since it is based on simple iterations with low computational complexity and has been successfully applied to numerous problems in various fields [24]-[28].

In order to describe the PSO algorithm, consider the minimization of an objective function over parameter θ . In PSO, first a number of parameter values $\{\theta_i\}_{i=1}^M$, called *particles*, are generated, where M is called the population size (i.e., the number of particles). Then, iterations are performed, where at each iteration new particles are generated as the summation of the previous particles and velocity vectors \mathbf{v}_i according to the following equations [12]:

$$\mathbf{v}_i^{k+1} = \chi \left(\omega \mathbf{v}_i^k + c_1 \rho_{i1}^k (\mathbf{p}_i^k - \theta_i^k) + c_2 \rho_{i2}^k (\mathbf{p}_g^k - \theta_i^k) \right) \quad (22)$$

$$\theta_i^{k+1} = \theta_i^k + \mathbf{v}_i^{k+1} \quad (23)$$

for $i = 1, \dots, M$, where k is the iteration index, χ is the constriction factor, ω is the inertia weight, which controls the effects of the previous history of velocities on the current velocity, c_1 and c_2 are the cognitive and social parameters, respectively, and ρ_{i1}^k and ρ_{i2}^k are independent uniformly distributed random variables on $[0, 1]$ [21]. In (22), \mathbf{p}_i^k represents the position corresponding to the smallest objective function value until the k th iteration of the i th particle, and \mathbf{p}_g^k denotes the position corresponding to the global minimum among all the particles until the k th iteration. After a number of iterations, the position with the lowest objective function value, \mathbf{p}_g^k , is selected as the optimizer of the optimization problem.

In order to extend PSO to constrained optimization problems, various approaches, such as penalty functions and keeping feasibility of particles, can be taken [22], [23]. In the penalty function approach, a particle that becomes infeasible is assigned a large value (considering a minimization problem), which forces migration of particles to the feasible region. In the constrained optimization approach that preserves the feasibility of the particles, no penalty is applied to any particles; but for the positions \mathbf{p}_i^k and \mathbf{p}_g^k in (22) corresponding to the lowest objective function values, only the feasible particles are considered [23].

In order to employ PSO for the optimal stochastic signaling problem in (21), the optimization variable is defined as $\theta \triangleq [x_1 \ x_2 \ x_3 \ \lambda_1 \ \lambda_2 \ \lambda_3]^T$, and the iterations in (22) and (23) are used while using a penalty function approach to impose the constraints. The results are presented in Section IV.

2) *Convex Optimization Approach:* In order to provide an alternative approximate solution with lower complexity, consider a scenario in which the PDF of the signal is modeled as

$$p_S(x) = \sum_{j=1}^K \tilde{\lambda}_j \delta(x - \tilde{x}_j), \quad (24)$$

where \tilde{x}_j 's are the *known* mass points of the PDFs, and $\tilde{\lambda}_j$'s are the weights to be estimated. This scenario corresponds to the cases with a finite number of possible signal values. For example, in a digital communications system, if the transmitter can only send one of K pre-determined \tilde{x}_j values for a specific symbol, then the problem becomes calculating the optimal probability assignments, $\tilde{\lambda}_j$'s, for the possible signal values for each symbol. Note that since the optimization is performed

over PDFs as in (24), the optimal solution can include more than three mass points in general. In other words, the solution in this case is expected to approximate the optimal PDF, which includes at most three mass points, with a PDF with multiple mass points.

The solution to the optimal signal design problem in (10) over the set of signals with their PDFs as in (24) can be obtained from the solution of the following convex optimization problem:⁵

$$\begin{aligned} \min_{\tilde{\lambda}} \quad & \mathbf{g}^T \tilde{\lambda} \\ \text{subject to} \quad & \mathbf{B} \tilde{\lambda} \preceq \mathbf{C}, \\ & \mathbf{1}^T \tilde{\lambda} = 1, \quad \tilde{\lambda} \succeq \mathbf{0}, \end{aligned} \quad (25)$$

where $\mathbf{g} \triangleq [G(\tilde{x}_1) \cdots G(\tilde{x}_K)]^T$, with $G(x)$ as in (9),

$$\mathbf{B} \triangleq \begin{bmatrix} \tilde{x}_1^2 & \cdots & \tilde{x}_K^2 \\ \tilde{x}_1^4 & \cdots & \tilde{x}_K^4 \end{bmatrix}, \quad \mathbf{C} \triangleq \begin{bmatrix} A \\ \kappa A^2 \end{bmatrix}, \quad (26)$$

and $\mathbf{1}$ and $\mathbf{0}$ represent vectors of all ones and all zeros, respectively.

It is observed from (25) that the optimal weight assignments can be obtained as the solution of a convex optimization problem, specifically, a linearly constrained linear programming problem. Therefore, the solution can be obtained in polynomial time [13].

Note that if the set of possible signal values \tilde{x}_j 's include the deterministic signal value for the conventional algorithm, i.e., \sqrt{A} , then the performance of the convex algorithm in (25) can never be worse than that of the conventional one. In addition, as the number of possible signal values, K in (24), increases, the convex algorithm can approximate the exact optimal solution more closely.

IV. SIMULATION RESULTS

In this section, numerical examples are presented for a binary communications system with equal priors ($\pi_0 = \pi_1 = 0.5$) in order to investigate the theoretical results in the previous section. In the implementation of the PSO algorithm specified by (22) and (23), $M = 50$ particles are employed and 10000 iterations are performed. In addition, the parameters are set to $c_1 = c_2 = 2.05$ and $\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number [12]. Also, a penalty function approach is implemented to impose the constraints in (21); namely, the objective function is set to 1 whenever a particle becomes infeasible [24].

First, the noise in (1) is modeled by the PDF in (19), $A = 0.64$ and $\kappa = 1.5$ are employed for the constraints in (10), and the decision rule at the receiver is specified by $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$ (that is, a sign detector). As stated after (20), the conventional signaling is suboptimal in this case based on Proposition 2. In order to calculate optimal signals via the PSO and the convex optimization algorithms in Section III-D, the optimization problems in (21) and (25) are solved, respectively. For the convex algorithm, the mass points \tilde{x}_j in (24) are selected uniformly over the interval $[0, 2]$ with a step size of Δ , and the results for $\Delta = 0.01$ and $\Delta = 0.1$ are considered. Fig. 1 illustrates the optimal probability distributions obtained from the PSO and the convex

⁵For K -dimensional vectors \mathbf{x} and \mathbf{y} , $\mathbf{x} \preceq \mathbf{y}$ means that the i th element of \mathbf{x} is smaller than or equal to the i th element of \mathbf{y} for $i = 1, \dots, K$.

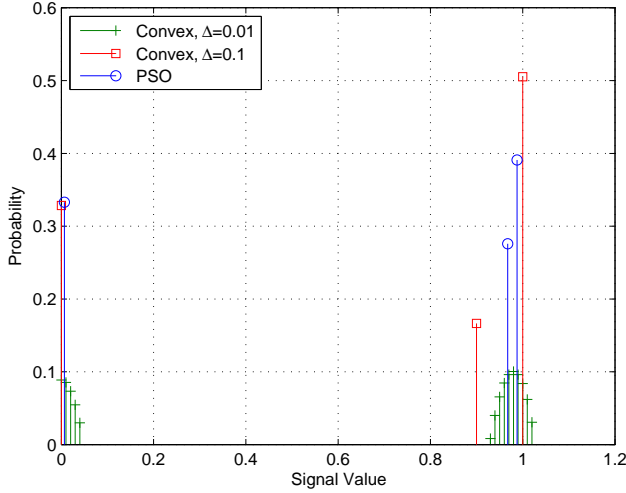


Fig. 1. Probability mass functions (PMFs) of the PSO and the convex optimization algorithms for the noise PDF in (19).

optimization algorithms.⁶ It is calculated that the conventional algorithm, which uses a deterministic signal value of 0.8, has an average error probability of 0.3293, whereas the PSO and the convex optimization algorithms with $\Delta = 0.01$ and $\Delta = 0.1$ have average error probabilities of 0.2909, 0.2911 and 0.2912, respectively. It is noted that the PSO algorithm achieves the lowest error probability with three mass points and the convex algorithms approximate the PSO solution with multiple mass points around those of the PSO solution. In addition, the calculations indicate that the optimal solutions achieve both the second and the fourth moment constraints in accordance with Proposition 4-b.

Next, the optimal signaling problem is studied in the presence of Gaussian mixture. The Gaussian mixture noise can be used to model the effects of co-channel interference, impulsive noise and multiuser interference in communications systems [5], [7]. In the simulations, the Gaussian mixture noise is specified by $p_N(y) = \sum_{l=1}^L v_l \psi_l(y - y_l)$, where $\psi_l(y) = e^{-y^2/(2\sigma_l^2)}/(\sqrt{2\pi}\sigma_l)$. In this case, $G(x)$ can be obtained from (9) as $G(x) = \sum_{l=1}^L v_l Q((x + y_l)/\sigma_l)$. In all the scenarios, the variance parameter for each mass point of the Gaussian mixture is set to σ^2 (i.e., $\sigma_l^2 = \sigma^2 \forall l$), and the average power constraint A is set to 1. Note that the average power of the noise can be calculated as $E\{N^2\} = \sigma^2 + \sum_{l=1}^L v_l y_l^2$. First, we consider a symmetric Gaussian mixture noise which has its mass points at $\pm[0.3 \ 0.455 \ 1.011]$ with corresponding weights $[0.1 \ 0.317 \ 0.083]$ in order to illustrate the improvements that can be obtained via stochastic signaling. In Fig. 2, the average error probabilities of various algorithms are plotted against A/σ^2 when $\kappa = 1.1$ for both the sign detector and the ML detector. For the sign detector, the decision rule at the receiver is specified by $\Gamma_0 = (-\infty, 0]$ and $\Gamma_1 = [0, \infty)$. In this case, it is observed from Fig. 2 that the conventional algorithm, which uses a constant signal value of 1, has a large error floor compared to the PSO and convex optimization

⁶For the probability distributions obtained from the convex optimization algorithms, the signal values that have zero probability are not marked in the figures to clarify the illustrations.

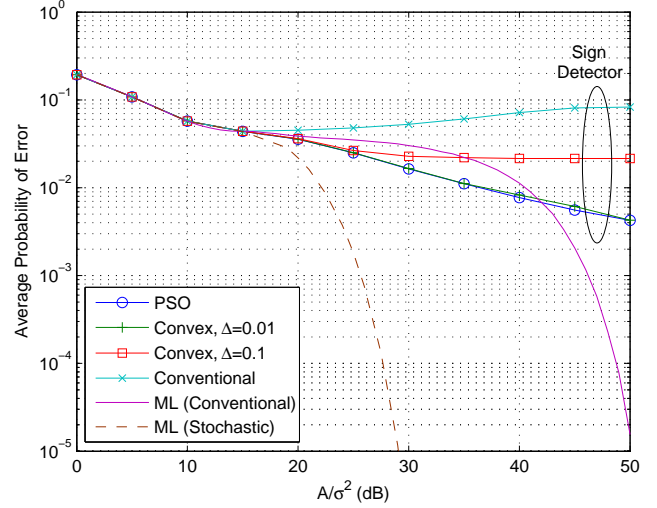


Fig. 2. Error probability versus A/σ^2 for $\kappa = 1.1$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3 \ 0.455 \ 1.011]$ with corresponding weights $[0.1 \ 0.317 \ 0.083]$, is considered.

algorithms at high A/σ^2 . Also, the average probability of error of the conventional signaling increases as A/σ^2 increases after a certain value. This seemingly counterintuitive result is observed because the average probability of error is related to the area under the two shifted noise PDFs as in (5). Since the noise has a multi-modal PDF, that area is a non-monotonic function of A/σ^2 and can increase in some cases as A/σ^2 increases. It is also observed that the convex optimization algorithm performs very closely to the PSO algorithm for densely spaced possible signal values, i.e., for $\Delta = 0.01$. For the ML detector, the receiver compares $p_N(y - \sqrt{A})$ and $p_N(y + \sqrt{A})$, and decides symbol 0 if the latter is larger, and decides 1 otherwise. It is observed for small σ^2 values that the ML receiver performs significantly better than the other receivers that are based on the sign detector. However, stochastic signaling causes the sign detector to perform better than the conventional ML receiver, which uses deterministic signaling, for medium A/σ^2 values. For example, the PSO and convex optimization algorithms for $\Delta = 0.01$ have better performance than the ML receiver for A/σ^2 values from 20 dB to 40 dB. This is mainly due to the fact that the conventional ML detector uses deterministic signaling whereas the others employ stochastic signaling. However, when the stochastic signaling is applied to the ML detector as well, it achieves the lowest probabilities of error for all A/σ^2 values as observed in Fig. 2 (labeled as “ML (Stochastic)”).

Another observation from Fig. 2 is that improvements over the conventional algorithm disappear as σ^2 increases (i.e., for small A/σ^2 values). This result can be explained from Propositions 1 and 2, based on the plots of $G(x)$ at various A/σ^2 values. For example, Fig. 3 illustrates the plots of $G(x)$ at A/σ^2 of 0, 20 and 40 dB for the sign detector. The function is decreasing and convex for 0 dB for the positive signal values, which are practically the domain of optimization since $G(x)$ is a decreasing function and the constraint functions x^2

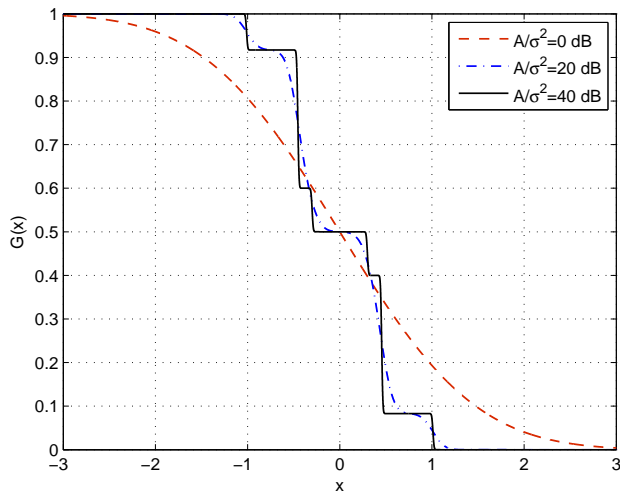


Fig. 3. $G(x)$ in (9) for the sign detector in Fig. 2 at A/σ^2 values of 0, 20 and 40 dB.

and x^4 are even functions.⁷ Therefore, Proposition 1 implies that the conventional algorithm that uses a constant signal value of 1 is optimal in this case, as observed in Fig. 2. On the other hand, at 20 dB and 40 dB, the calculations show that the condition in Proposition 2 is satisfied; hence, the conventional algorithm cannot be optimal in that case, and improvements are observed in Fig. 2 at $A/\sigma^2 = 20$ dB and $A/\sigma^2 = 40$ dB. Another result obtained from the numerical studies for Fig. 2 is that all the solutions achieve at least one of the second moment or the fourth moment constraints with equality as a result of Proposition 4.

For the scenario in Fig. 2, the probability distributions of the optimal signals for the sign detector are shown in Fig. 4 and Fig. 5 for $A/\sigma^2 = 20$ dB and $A/\sigma^2 = 40$ dB, respectively, where both the PSO and the convex optimization algorithms are considered. In the first case, the convex optimization algorithm with $\Delta = 0.1$ approximates the probability mass function (PMF) obtained from the PSO algorithm with two mass points (with nonzero probabilities), whereas the convex optimization algorithm with $\Delta = 0.01$ results in 8 mass points. In the second case, the convex optimization algorithms with $\Delta = 0.1$ and $\Delta = 0.01$ result in PMFs with two and three mass points, respectively, as shown in Fig. 5. Since the convex optimization algorithm with $\Delta = 0.1$ does not provide a PMF that is very close to those of the other algorithms in this case, the resulting error probability becomes significantly higher for that algorithm, as observed from Fig. 2 at $A/\sigma^2 = 40$ dB.

Finally, a symmetric Gaussian mixture noise which has its mass points at $\pm[0.19 \ 0.39 \ 0.83 \ 1.03]$ each with a weight of $1/8$ is considered. Such a noise PDF can be considered to model the effects of co-channel interference [7], or a system that operates under the effect of multiuser interference [5]. For example, in the presence of multiple users, the noise can be modeled as $N = \sum_{k=2}^K A_k b_k + \eta$, where $b_k \in \{-1, 1\}$ with equal probabilities and η is a zero-mean Gaussian thermal noise component with variance σ^2 . Then, for $K = 4$, $A_2 =$

⁷In other words, negative signal values are never selected for symbol 1 since selecting the absolute value of a negative signal value always gives a smaller average probability of error without changing the signal moments.

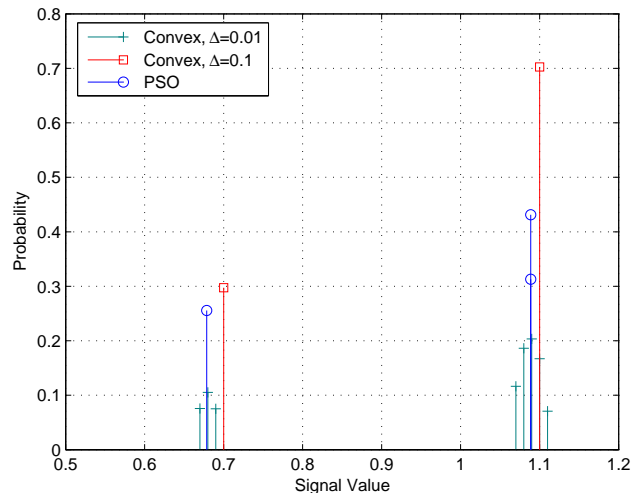


Fig. 4. PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2 at $A/\sigma^2 = 20$ dB.

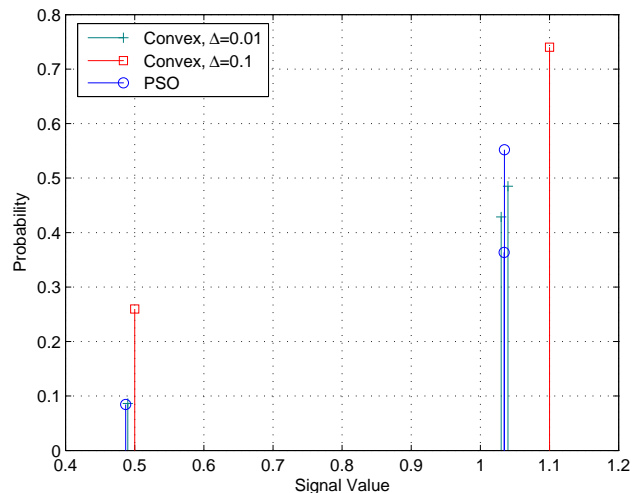


Fig. 5. PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 2 at $A/\sigma^2 = 40$ dB.

0.1 , $A_3 = 0.61$ and $A_2 = 0.32$, the noise becomes Gaussian mixture noise with 8 mass points as specified at the beginning of the paragraph. In Fig. 6, the average error probabilities of various algorithms are plotted against the A/σ^2 for $\kappa = 1.5$. Also the plots of $G(x)$ at $A/\sigma^2 = 0, 25, 40$ dB are presented in Fig. 7, and the probability distributions at $A/\sigma^2 = 25$ dB and $A/\sigma^2 = 40$ dB are illustrated in Fig. 8 and Fig. 9, respectively, for the sign detector. Although similar observations as in the previous scenario can be made, a number of differences are also noticed. The improvements achieved via the stochastic signaling over the conventional (deterministic) signaling are less than those observed in Fig. 2. In addition, since $\kappa = 1.5$ in this scenario, only the second moment constraint is achieved with equality in all the solutions.

In order to investigate the optimal stochastic signaling for the ML detectors studied in Fig. 2 and Fig. 6, Table I presents the PDFs of the optimal stochastic signals in those scenarios,

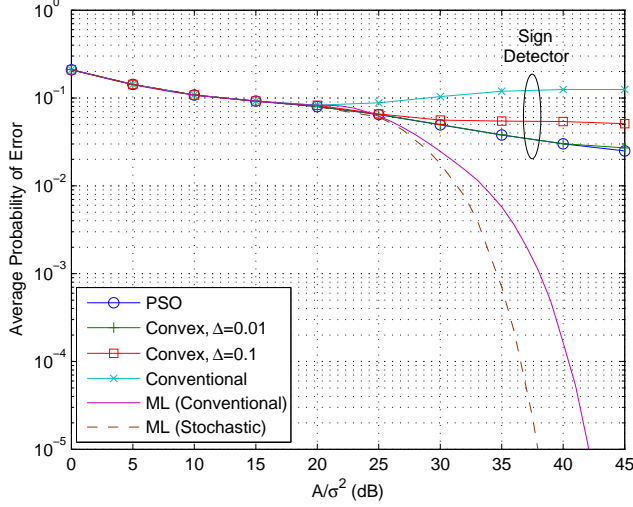


Fig. 6. Error probability versus A/σ^2 for $\kappa = 1.5$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.19 \ 0.39 \ 0.83 \ 1.03]$, each with equal weight, is considered.

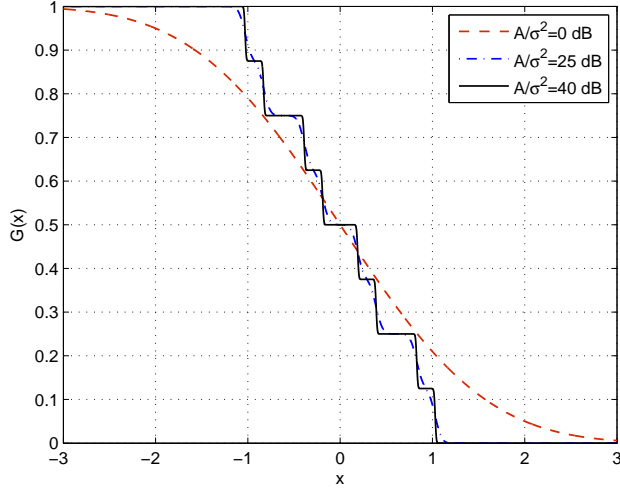


Fig. 7. $G(x)$ in (9) for the sign detector in Fig. 6 at A/σ^2 values of 0, 25 and 40 dB.

TABLE I

OPTIMAL STOCHASTIC SIGNALS FOR THE ML DETECTORS IN FIG. 2 (TOP BLOCK) AND FIG. 6 (BOTTOM BLOCK).

A/σ^2 (dB)	λ_1	λ_2	λ_3	x_1	x_2	x_3
10	1	0	0	1	N/A	N/A
15	1	0	0	1	N/A	N/A
20	0.1181	0.8819	0	1.4211	0.9151	N/A
25	0.1264	0.8736	0	1.4494	0.8876	N/A
27.5	0.1317	0.8683	0	1.4465	0.8811	N/A
10	1	0	0	1	N/A	N/A
15	1	0	0	1	N/A	N/A
20	0.1272	0.8728	0	0.5073	1.0527	N/A
25	0.9791	0.0209	0	0.9950	1.2116	N/A
30	0.9415	0.0585	0	0.9859	1.2047	N/A
35	0.9236	0.0764	0	0.9823	1.1936	N/A

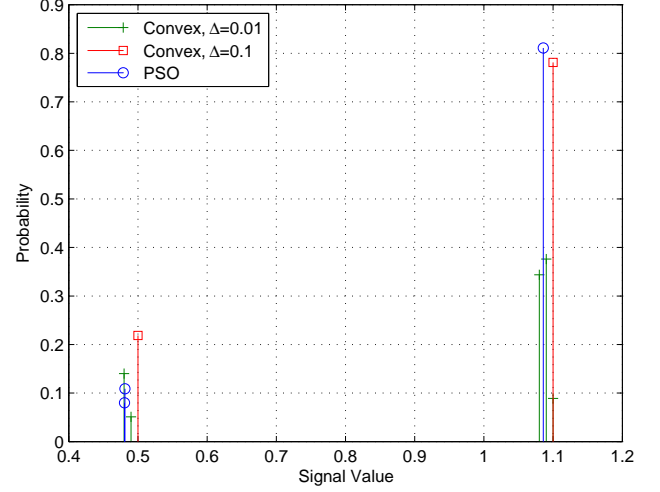


Fig. 8. PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 6 at $A/\sigma^2 = 25$ dB.

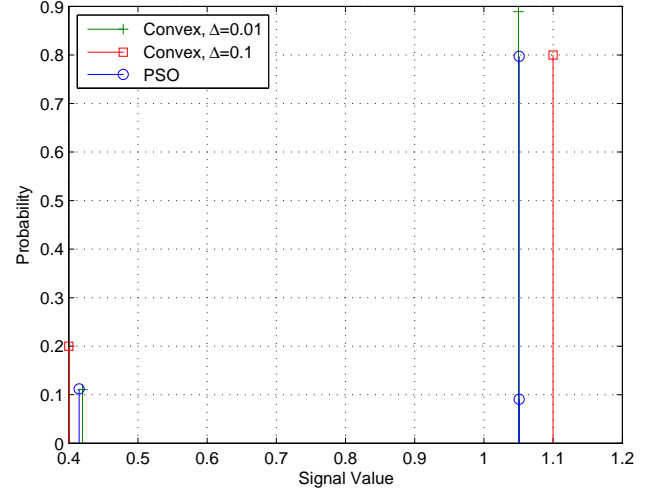


Fig. 9. PMFs of the PSO and the convex optimization algorithms for the sign detector in Fig. 6 at $A/\sigma^2 = 40$ dB.

where the optimal PDFs are expressed in the form of $p_S(x) = \lambda_1 \delta(x - x_1) + \lambda_2 \delta(x - x_2) + \lambda_3 \delta(x - x_3)$. It is observed from the table that the conventional deterministic signaling is optimal at low A/σ^2 values, which can also be verified from Fig. 2 and Fig. 6 since there is no improvement via the stochastic signaling over the conventional one for those A/σ^2 values. However, as A/σ^2 increases, the optimal signaling is achieved via randomization between two signal values. In those cases, significant improvements over the conventional signaling can be achieved as observed from Fig. 2 and Fig. 6. Finally, it is noted from the table that the optimal solutions result in randomization between at most two different signal levels in this example. This is in compliance with Proposition 3 since the proposition does not guarantee the existence of three different signal levels in general but states that an optimal signal can be represented by a randomization of *at most* three different signal levels.

V. EXTENSIONS TO M -ARY PULSE AMPLITUDE MODULATION (PAM)

The results in the study can be extended to M -ary PAM communications systems for $M > 2$ as well. To that aim, consider a generic detector which chooses the i th symbol if the observation is in decision region Γ_i for $i = 0, 1, \dots, M-1$. In other words, the decision rule is defined as

$$\phi(y) = i, \quad \text{if } y \in \Gamma_i, \quad i = 0, 1, \dots, M-1. \quad (27)$$

Then, the average probability of error for an M -ary system can be expressed as

$$P_{\text{avg}} = \sum_{i=0}^{M-1} \pi_i (1 - P_i(\Gamma_i)), \quad (28)$$

where π_i denotes the prior probability of the i th symbol.

If signals S_0, S_1, \dots, S_{M-1} are modeled as stochastic signals with PDFs $p_{S_0}, p_{S_1}, \dots, p_{S_{M-1}}$, respectively, the average probability of error in (28) can be expressed, similarly to (6), as

$$P_{\text{avg}}^{\text{stoc}} = \sum_{i=0}^{M-1} \pi_i \left(1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t) dy dt \right). \quad (29)$$

Then, the optimal stochastic signaling problem can be stated as

$$\begin{aligned} \min_{p_{S_0}, \dots, p_{S_{M-1}}} & \sum_{i=0}^{M-1} \pi_i \left(1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t) dy dt \right) \\ \text{subject to} & \quad E\{|S_i|^2\} \leq A, \quad E\{|S_i|^4\} \leq \kappa A^2, \\ & \quad i = 0, 1, \dots, M-1. \end{aligned} \quad (30)$$

Due to the structure of the objective function in (30) and the individual constraints on each signal, M separate optimization problems, similar to (8), can be obtained. Namely, for $i = 0, 1, \dots, M-1$,

$$\begin{aligned} \min_{p_{S_i}} & 1 - \int_{-\infty}^{\infty} p_{S_i}(t) \int_{\Gamma_i} p_N(y-t) dy dt \\ \text{subject to} & \quad E\{|S_i|^2\} \leq A, \quad E\{|S_i|^4\} \leq \kappa A^2. \end{aligned} \quad (31)$$

In addition, if auxiliary functions $G_i(x)$ are defined as $G_i(x) \triangleq 1 - \int_{\Gamma_i} p_N(y-x) dy$ for $i = 0, 1, \dots, M-1$, the optimization problem in (31) can be expressed as

$$\begin{aligned} \min_{p_{S_i}} & E\{G_i(S_i)\} \\ \text{subject to} & \quad E\{|S_i|^2\} \leq A, \quad E\{|S_i|^4\} \leq \kappa A^2 \end{aligned} \quad (32)$$

for $i = 0, 1, \dots, M-1$. Since (32) is in the same form as (10), the results in Section III can be extended to M -ary PAM systems, as well.

VI. CONCLUDING REMARKS AND EXTENSIONS

In this paper, the stochastic signaling problem under second and fourth moment constraints has been studied for binary communications systems. It has been shown that, under certain monotonicity and convexity conditions, the conventional signaling, which employs deterministic signals at the average power limit, is optimal. On the other hand, in some cases, a smaller average probability of error can be achieved by using a

signal that is obtained by a randomization of multiple signal values. In addition, it has been shown that an optimal signal can be represented by a discrete random variable with at most three mass points, which simplifies the optimization problem for the optimal signal design considerably. Furthermore, it has been observed that the optimal signals achieve at least one of the second and fourth moment constraints in most practical scenarios. Finally, two techniques based on PSO and convex relaxation have been proposed to obtain the optimal signals, and simulation results have been presented.

In addition, the results in this paper can be extended to a generic binary hypothesis-testing problem in the Bayesian framework [2], [14].⁸ In that case, the average probability of error expression in (3) is generalized to the *Bayes risk*, defined as $\pi_0[C_{00}P_0(\Gamma_0) + C_{10}P_0(\Gamma_1)] + \pi_1[C_{01}P_1(\Gamma_0) + C_{11}P_1(\Gamma_1)]$, where $C_{ij} \geq 0$ represents the cost of deciding the i th hypothesis when the j th one is true. Then, all the results in the paper are still valid when function G in (9) is replaced by $G(x) = C_{01} \int_{\Gamma_0} p_N(y-x) dy + C_{11} \int_{\Gamma_1} p_N(y-x) dy$. Moreover, it can be shown that the results in this paper are valid in the *minimax* and *Neyman-Pearson* frameworks [2] due to the decoupling of the optimization problem discussed in Section III.

APPENDIX

A. Derivation of (8)

The optimal stochastic signaling problem in (7) can be expressed from (6) as

$$\begin{aligned} \min_{p_{S_0}, p_{S_1}} & \pi_0 \int_{-\infty}^{\infty} p_{S_0}(t) \int_{\Gamma_1} p_N(y-t) dy dt \\ & + \pi_1 \int_{-\infty}^{\infty} p_{S_1}(t) \int_{\Gamma_0} p_N(y-t) dy dt \end{aligned} \quad (33)$$

$$\text{subject to} \quad E\{|S_0|^2\} \leq A, \quad E\{|S_0|^4\} \leq \kappa A^2 \quad (34)$$

$$E\{|S_1|^2\} \leq A, \quad E\{|S_1|^4\} \leq \kappa A^2 \quad (35)$$

For a given decision rule (detector) and a noise PDF, changing p_{S_0} has no effect on the second term in (33) and the constraints in (35). Similarly, changing p_{S_1} has no effect on the first term in (33) and the constraints in (34). Therefore, the problem of minimizing the expression in (33) over p_{S_0} and p_{S_1} under the constraints in (34) and (35) is equivalent to minimizing the first term in (33) over p_{S_0} under the constraints in (34) and minimizing the second term in (33) over p_{S_1} under the constraints in (35). Therefore, the signal design problems for S_0 and S_1 can be separated as in (8). \square

B. Proof of Proposition 3

In order to prove Proposition 3, we take an approach similar to those in [11] and [29]. First, the following set is defined:

$$U = \{(u_1, u_2, u_3) : u_1 = G(x), u_2 = x^2, u_3 = x^4, \text{ for } |x| \leq \gamma\}. \quad (36)$$

Since $G(x)$ is continuous, the mapping from $[-\gamma, \gamma]$ to \mathbb{R}^3 defined by $F(x) = (G(x), x^2, x^4)$ is continuous. Since the continuous image of a compact set is compact, U is a compact set [30].

⁸Hence, the results in the paper can be applied to other systems than communications, as well.

Let V represent the convex hull of U . Since U is compact, the convex hull V of U is closed [30]. Also, the dimension of V should be smaller than or equal to 3, since $V \subseteq \mathbb{R}^3$. In addition, let W be the set of all possible conditional error probability $P_1(\Gamma_0)$, second moment, and fourth moment triples; i.e.,

$$W = \left\{ (w_1, w_2, w_3) : w_1 = \int_{-\infty}^{\infty} p_S(x)G(x)dx, \right. \\ w_2 = \int_{-\infty}^{\infty} p_S(x)x^2dx, w_3 = \int_{-\infty}^{\infty} p_S(x)x^4dx, \\ \left. \forall p_S(x), |x| \leq \gamma \right\}, \quad (37)$$

where $p_S(x)$ is the signal PDF.

Similar to [29], $V \subseteq W$ can be proven as follows. Since V is the convex hull of U , each element of V can be expressed as $\mathbf{v} = \sum_{i=1}^L \lambda_i (G(x_i), x_i^2, x_i^4)$, where $\sum_{i=1}^L \lambda_i = 1$, and $\lambda_i \geq 0 \forall i$. Considering set W , it has an element that is equal to \mathbf{v} for $p_S(x) = \sum_{i=1}^L \lambda_i \delta(x - x_i)$. Hence, each element of V also exists in W . On the other hand, since for any vector random variable Θ that takes values in set Ω , its expected value $E\{\Theta\}$ is in the convex hull of Ω [11], it is concluded from (36) and (37) that W is in the convex hull V of U ; that is, $V \supseteq W$ [31].

Since $W \supseteq V$ and $V \supseteq W$, it is concluded that $W = V$. Therefore, Carathéodory's theorem [32], [33] implies that any point in V (hence, in W) can be expressed as the convex combination of at most 4 points in U . Since an optimal PDF should minimize the average probability of error, it corresponds to the boundary of V . Since V is a closed set as discussed at the beginning of the proof, it contains its own boundary. Since any point at the boundary of V can be expressed as the convex combination of at most 3 elements in U [32], an optimal PDF can be represented by a discrete random variable with 3 mass points.

C. Proof of Proposition 4

a) Let $A^2 < x_{\min}^4 < \kappa A^2$ and $p_{S_1}(x)$ represent an optimal signal PDF with $w_1 \triangleq E\{G(S)\}$, $w_2 \triangleq E\{S^2\}$ and $w_3 \triangleq E\{S^4\}$, where $w_2 < A$ and $w_3 \leq \kappa A^2$. In the following, it is shown that such a signal cannot be optimal (hence, a contradiction), and an optimal signal needs to satisfy $E\{S^2\} = A$. To that aim, define another signal PDF as follows:

$$p_{S_2}(x) = \frac{A - w_2}{x_{\min}^2 - w_2} \delta(x - x_{\min}) + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} p_{S_1}(x). \quad (38)$$

It can be shown for $p_{S_2}(x)$ that

$$E\{G(S)\} = \frac{A - w_2}{x_{\min}^2 - w_2} G(x_{\min}) + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} w_1 < w_1, \quad (39)$$

$$E\{S^2\} = \frac{A - w_2}{x_{\min}^2 - w_2} x_{\min}^2 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} w_2 = A, \quad (40)$$

$$E\{S^4\} = \frac{A - w_2}{x_{\min}^2 - w_2} x_{\min}^4 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} w_3 < \kappa A^2. \quad (41)$$

The inequality in (39) is obtained by observing that $G(x_{\min})$ is the unique minimum value of $G(x)$ and that no signals can achieve $E\{G(S)\} = G(x_{\min})$ since $x_{\min} > \sqrt{A}$. The inequality in (41) is achieved since $x_{\min}^4 < \kappa A^2$ and $w_3 \leq \kappa A^2$. From (39)-(41), it is concluded that $p_{S_2}(x)$ defines a better signal than $p_{S_1}(x)$ does. In other words, the optimal signal cannot have a smaller average power than A ; that is, $E\{S^2\} = A$ must be satisfied by the optimal signal.

b) Now assume $x_{\min}^4 > \kappa A^2$ and $p_{S_1}(x)$ represents an optimal signal PDF with $w_1 \triangleq E\{G(S)\}$, $w_2 \triangleq E\{S^2\}$ and $w_3 \triangleq E\{S^4\}$, where $w_2 < A$ and $w_3 < \kappa A^2$. In the following, it is proven that $w_2 < A$ and $w_3 < \kappa A^2$ cannot be satisfied at the same time for an optimal signal.

Consider $p_{S_2}(x)$ in (38) and $p_{S_3}(x)$ below:

$$p_{S_3}(x) = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3} \delta(x - x_{\min}) + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3} p_{S_1}(x). \quad (42)$$

For both $p_{S_2}(x)$ and $p_{S_3}(x)$, it can be shown that $E\{G(S)\} < w_1$ since $G(x_{\min}) < w_1$. For $p_{S_2}(x)$, the second and fourth moment constraints can be expressed as

$$E\{S^2\} = \frac{A - w_2}{x_{\min}^2 - w_2} x_{\min}^2 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} w_2 = A, \quad (43)$$

$$E\{S^4\} = \frac{A - w_2}{x_{\min}^2 - w_2} x_{\min}^4 + \frac{x_{\min}^2 - A}{x_{\min}^2 - w_2} w_3 \triangleq \beta_1. \quad (44)$$

On the other hand, for $p_{S_3}(x)$, the constraints are given by

$$E\{S^2\} = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3} x_{\min}^2 + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3} w_2 \triangleq \beta_2, \quad (45)$$

$$E\{S^4\} = \frac{\kappa A^2 - w_3}{x_{\min}^4 - w_3} x_{\min}^4 + \frac{x_{\min}^4 - \kappa A^2}{x_{\min}^4 - w_3} w_3 = \kappa A^2. \quad (46)$$

Now it is claimed that at least one of the conditions $\beta_1 \leq \kappa A^2$ or $\beta_2 \leq A$ must be true. In other words, it is not possible to have $\beta_1 > \kappa A^2$ and $\beta_2 > A$ at the same time. To prove this, the condition for $\beta_1 > \kappa A^2$ is considered first. Since $x_{\min}^4 > \kappa A^2$ and $w_3 < \kappa A^2$, $\beta_1 > \kappa A^2$ can be expressed from (44) as

$$\frac{x_{\min}^4 - \kappa A^2}{\kappa A^2 - w_3} > \frac{x_{\min}^2 - A}{A - w_2}. \quad (47)$$

Next, the $\beta_2 > A$ condition is considered. Since $x_{\min}^2 > A$ and $w_2 < A$, that condition can be expressed, from (45), as

$$\frac{x_{\min}^4 - \kappa A^2}{\kappa A^2 - w_3} < \frac{x_{\min}^2 - A}{A - w_2}. \quad (48)$$

Since (47) and (48) cannot be true at the same time, at least one of the conditions $\beta_1 \leq \kappa A^2$ or $\beta_2 \leq A$ is true. This implies that at least one of $p_{S_2}(x)$ or $p_{S_3}(x)$ provides a signal that has a smaller average probability of error than that for $p_{S_1}(x)$. In addition, such a signal satisfies at least one of the constraints with equality as can be observed from (43) and (46). Therefore, an optimal signal cannot be in the form of $p_{S_1}(x)$, which satisfies both inequalities as $E\{S^2\} < A$ and $E\{S^4\} < \kappa A^2$.

Acknowledgments: The authors would like to thank Prof. Erdal Arkan and Suat Bayram from Bilkent University, Prof. Baris Coskunuzer from Koc University, and Prof. Mustafa Cenk Gursoy from the University of Nebraska–Lincoln for their insightful comments.

REFERENCES

- [1] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [2] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
- [3] I. Korn, J. P. Fonseka, and S. Xing, "Optimal binary communication with nonequal probabilities," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1435–1438, Sep. 2003.
- [4] F. Cabarcas, R. D. Souza, and J. Garcia-Frias, "Turbo coding of strongly nonuniform memoryless sources with unequal energy allocation and PAM signaling," *IEEE Trans. Sig. Processing*, vol. 54, no. 5, pp. 1942–1946, May 2006.
- [5] S. Verdú, *Multuser Detection*. 1st ed. Cambridge, UK: Cambridge University Press, 1998.
- [6] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.
- [7] V. Bhatia and B. Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Processing*, vol. 87, pp. 2569–2586, Nov. 2007.
- [8] S. Shamaï and S. Verdú, "Worst-case power-constrained noise for binary-input channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1494–1511, Sep. 1992.
- [9] M. A. Klimesh and W. E. Stark, "Worst-case power-constrained noise for binary-input channels with varying amplitude signals," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, July 1994, p. 381.
- [10] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.
- [11] L. Huang and M. J. Neely, "The optimality of two prices: Maximizing revenue in a stochastic network," in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2007.
- [12] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*. IOS Press, 2002, pp. 214–220, in *Intelligent Technologies—Theory and Applications: New Trends in Intelligent Technologies*.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice Hall, Inc., 1998.
- [15] A. L. McKellips and S. Verdú, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Trans. Inform. Theory*, vol. 43, no. 4, pp. 1256–1264, July 1997.
- [16] T. Erseghe, V. Cellini, and G. Dona, "On UWB impulse radio receivers derived by modeling MAI as a Gaussian mixture process," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2388–2396, June 2008.
- [17] M. C. Gursoy, H. V. Poor, and S. Verdú, "Efficient signaling for low-power rician fading channels," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2002.
- [18] M. McGuire, "Location of mobile terminals with quantized measurements," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Commun. (PIMRC)*, vol. 3, Berlin, Germany, Sep. 2005, pp. 2045–2049.
- [19] T. Erseghe and S. Tomasin, "Optimized demodulation for MAI resilient UWB W-PAN receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 4867–4871.
- [20] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.
- [21] A. I. F. Vaz and E. M. G. P. Fernandes, "Optimization of nonlinear constrained particle swarm," *Baltic Journal on Sustainability*, vol. 12, no. 1, pp. 30–36, 2006.
- [22] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evolutionary Computation*, vol. 7, no. 1, pp. 19–44, 1999.
- [23] X. Hu and R. Eberhart, "Solving constrained nonlinear optimization problems with particle swarm optimization," in *Proc. Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002)*, Orlando, FL, 2002.
- [24] Y. Chen and V. K. Dubey, "Ultrawideband source localization using a particle-swarm-optimized Capon estimator," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, Seoul, Korea, May 2005, pp. 2825–2829.
- [25] Y. Rahmat-Samii, D. Gies, and J. Robinson, "Particle swarm optimization (PSO): A novel paradigm for antenna designs," *The Radio Science Bulletin*, vol. 305, pp. 14–22, Sep. 2003.
- [26] Z. Yangyang, J. Chunlin, Y. Ping, L. Manlin, W. Chaojin, and W. Guangxing, "Particle swarm optimization for base station placement in mobile communication," in *Proc. IEEE International Conference on Networking, Sensing and Control*, vol. 1, May 2004, pp. 428–432.
- [27] W. Jatmiko, K. Sekiyama, and T. Fukuda, "A PSO-based mobile sensor network for odor source localization in dynamic environment: Theory, simulation and measurement," in *Proc. IEEE Congress on Evolutionary Computation*, Vancouver, BC, July 2006, pp. 1036–1043.
- [28] J. Pugh, A. Martinoli, and Y. Zhang, "Particle swarm optimization for unsupervised robotic learning," in *Proc. Swarm Intelligence Symposium (SIS)*, Pasadena, California, June 2005, pp. 92–99.
- [29] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors," *IEEE Trans. Sig. Processing*, vol. 55, no. 7, pp. 3172–3184, July 2007.
- [30] C. C. Pugh, *Real Mathematical Analysis*. New York: Springer-Verlag, 2002.
- [31] S. Bayram and S. Gezici, "Noise-enhanced M -ary hypothesis-testing in the minimax framework," in *Proc. International Conference on Signal Processing and Commun. Systems*, Omaha, Nebraska, Sep. 2009, pp. 31–36.
- [32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1968.
- [33] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Boston, MA: Athena Specific, 2003.



Cagri Goken (S'10) received the B.S. degree from Bilkent University, Ankara, Turkey in 2009. He is currently working towards the M.S. degree in the Department of Electrical and Electronics Engineering, Bilkent University. His research interests are in the fields of wireless communications and statistical signal processing. Currently, he has particular interest in stochastic signaling for communications systems.



Sinan Gezici (S'03, M'06) received the B.S. degree from Bilkent University, Turkey in 2001, and the Ph.D. degree in Electrical Engineering from Princeton University in 2006. From April 2006 to January 2007, he worked as a Visiting Member of Technical Staff at Mitsubishi Electric Research Laboratories, Cambridge, MA. Since February 2007, he has been an Assistant Professor in the Department of Electrical and Electronics Engineering at Bilkent University.

Dr. Gezici's research interests are in the areas of signal detection, estimation and optimization theory, and their applications to wireless communications and localization systems. Among his publications in these areas is the book *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols* (Cambridge University Press, 2008).



Orhan Arikan (M'91) was born in 1964 in Manisa, Turkey. In 1986, he received his B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey. He received both his M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign, in 1988 and 1990, respectively.

Following his graduate studies, he was employed as a Research Scientist at Schlumberger-Doll Research Center, Ridgefield, CT. In 1993 he joined the Electrical and Electronics Engineering Department of Bilkent University, Ankara, Turkey. Since 2006, he is a full professor at Bilkent. His current research interests include statistical signal processing, time-frequency analysis and remote sensing. Dr. Arikan has served as chairman of Turkey chapter of IEEE Signal Processing Society and president of IEEE Turkey Section.