Optimal Stochastic Signaling under Average Power and Bit Rate Constraints

Cagri Goken, *Student Member, IEEE*, Berkan Dulek, and Sinan Gezici, *Senior Member, IEEE*

Abstract—The optimal stochastic signaling based on the joint design of prior distribution and signal constellation is investigated under average bit rate and power constraints. First, an optimization problem is formulated to maximize the average probability of correct decision over the set of joint distribution functions for prior probabilities and the corresponding constellation symbols. Next, an alternative problem formulation, for which the optimal joint distribution is characterized by a randomization among at most three mass points, is provided, and it is shown that both formulations share the same solution. Three special cases of the problem are investigated in detail. First, in the absence of randomization, the optimal prior probability distribution is analyzed for a given signal constellation and a closed-form solution is provided. Second, the optimal deterministic pair of prior probabilities and the corresponding signal levels is considered. Third, a binary communication system with scalar observations is investigated in the presence of a zero-mean additive white Gaussian noise, and the optimal solution is obtained under practical assumptions. Finally, numerical examples are presented to illustrate the theoretical results. It is observed that the proposed approach can provide improvements in terms of average symbol error rate over the classical scheme for certain scenarios.

Index Terms—Stochastic signaling, probability of error, prior probability, bit rate, power constraint.

I. INTRODUCTION AND MOTIVATION

In the literature, optimal signaling to minimize the average probability of error under various forms of power constraints has been studied extensively. For binary communication systems that operate over zero-mean additive white Gaussian noise (AWGN) channels subject to power constraints in the form of $\mathbb{E}\left\{\|\mathbf{S}_i\|^2\right\} \leq A$ for $i = 0, 1$, the optimal strategy is to employ deterministic antipodal signaling at the power limit at the transmitter and the maximum a *posteriori* probability (MAP) decision rule at the receiver [2]. Alternatively, the average power constraint can take the form of $\sum_{i=1}^{2} \pi_i \mathbb{E}\left\{\|\mathbf{S}_i\|^2\right\} \leq A$ where π_i represents prior probability of symbol i. In [3], the optimal deterministic signaling with such a constraint is investigated in the presence of additive zero-mean Gaussian noise when the optimal MAP receiver is used at the receiver, and it is shown for coherent systems that the optimum performance is achieved when the Euclidean distance between the signals is maximized under the given

power constraint and nonequal prior probabilities. In [4], the convexity properties of the average probability of error in terms of signal and noise power are investigated for binary-valued scalar signals over additive noise channels under an average power constraint. In [5], similar convexity analyses are performed for constellations with arbitrary shape, order, and dimensionality for a maximum likelihood (ML) detector in an AWGN channel. Based on the convexity results in [4] and [5], the optimality of deterministic or stochastic signaling can be determined in power constrained digital communication systems.

The problem of optimal constellation design (signal shaping) is also considered in various studies in the literature such as [6]–[12]. In [6], optimal nonuniform constellations to minimize the union bound on the uncoded symbol error rate are investigated in a cooperative relaying scheme. In [7], a nonuniform constellation design is performed to maximize the bit interleaved coded modulation (BICM) capacity for the ATSC 3.0 standard. The optimal two dimensional signal constellation which minimizes the probability of error over a circularly symmetric complex AWGN channel under average power constraints is investigated for M-ary communication systems in [8]. In [10], a nonequiprobable signaling scheme is described to achieve the asymptotic shaping gain (1.53 dB) in any fixed dimension.

In certain scenarios, employing randomization (i.e., stochastic signaling) instead of deterministic signals/constellation points can improve the average probability of error performance [4], [13]–[20]. Stochastic signaling relies on the idea of modeling signal $\mathbf{S}_i$ corresponding to the ith information symbol as a random variable instead of a deterministic quantity for each i. In [17], the optimal stochastic signaling is investigated for a given detector under second and fourth moment constraints, and it is shown that the optimal signal for each information symbol can be represented by a discrete random variable with at most three distinct signal levels. In [18], the joint design of the signals and the detector is investigated, and performance improvements over deterministic signaling are illustrated for non-Gaussian channels. In [19], optimal stochastic signaling is studied under an average power constraint in the form of $\sum_{i=1}^{2} \pi_i \mathbb{E}\left\{\|\mathbf{S}_i\|^2\right\} \leq A$ for $i = 0, 1$, and sufficient conditions for improvability or non-improvability of the deterministic signaling scheme given in [3] via stochastic signaling are derived. In [20], the stochastic signaling idea is applied in a downlink multiuser communication system. In particular, the optimal power control scheme is developed such that each user is allowed to randomize among multiple signal constellations instead of employing a fixed signal constella-

C. Goken and S. Gezici are with the Dept. of Electrical and Electronics Engineering, Bilkent University, 06800, Ankara, Turkey, Tel: +90-312-290-3139, Fax: +90-312-266-4192, emails: {cgoken,gezici}@ee.bilkent.edu.tr. B. Dulek is with the Dept. of Electrical and Electronics Engineering, Hacettepe University, 06800, Ankara, Turkey, email: berkan@ee.hacettepe.edu.tr.

Part of this work was presented at the 17th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2017), Bilbao, Spain, Dec. 2017 [1].

tion, and it is shown that randomization can improve error performance in some scenarios.

Although the optimal signaling has been investigated for a variety of power constraints and transmission scenarios in the literature, the prior probabilities are considered as fixed quantities, which can be either uniform or non-uniform. In conventional memoryless digital modulation systems, a uniform Bernoulli binary sequence is parsed into blocks of fixed length and each block is mapped to a symbol in a given signal constellation. Resulting in equally likely symbols, this procedure (i.e., uniform signaling) maximizes the entropy of the transmitted symbols, and consequently the average bit rate for a given constellation size [21]. In cases where the power cost of the constellation points also needs to be taken into account, a nonuniform signaling scheme that selects the constellation points with lower power more frequently than the points with higher power would result in power savings in exchange for a reduced bit rate [22]. In addition, it is known that for a given fixed signaling scheme, the minimum Bayesian risk (probability of error) is concave over the space of priors [2]. For example, for a binary communication system employing antipodal signaling ($\mathbf{S}_1 = -\mathbf{S}_0$), uniform priors result in the worst average symbol error rate. Therefore, non-uniform signaling can provide improvements for average error performance in addition to power savings even though it reduces the average bit rate.

Motivated by these observations, we consider the optimal signaling problem based on the joint design of prior probabilities and the corresponding constellation symbols such that the average symbol error rate is minimized under average bit rate and power constraints. To maintain a general perspective/formulation, both the prior probability vector and the signal constellation are assumed to be random (stochastic) distributed according to a joint probability density function (PDF), $p_{\mathbf{\Pi},\mathbf{S}}(\boldsymbol{\pi}, \mathbf{s})$. In other words, the transmitter forms an optimal constellation book in order to transmit each symbol with the corresponding signal levels and the prior probabilities, where each constellation can be used with a certain probability. This procedure can be regarded as a generalization of *constellation randomization*. In the literature, there exist some studies that utilize randomized signal constellations in various communication scenarios [23]–[27]. For example, in [23], for a spatial multiplexing scenario under block fading channels, the signal constellation is rotated by using a pseudorandom sequence for each transmitted vector. Performance gains via randomized constellations can be obtained both in coded frame-error rate [23] and outage probability [24]. In [25]–[27], random rotations and phase shifts are employed to increase the transmission diversity. Also, in [20], the optimal randomization of constellations is investigated for each user in a multiuser setting under power constraints. However, these studies do not take into account the prior probability distribution in their formulation (i.e., assume that it is fixed), and only utilize randomization in signal levels to achieve improvement according to a certain performance criterion.

In this paper, we consider an M-ary communication system with n dimensional observations. Our goal is to obtain the optimal joint distribution of the constellation symbols and

the corresponding prior probabilities to minimize the average probability of symbol error under average bit rate and power constraints. First, an optimization problem is formulated, where the receiver utilizes the optimal MAP decision rule by assuming that it knows the prior probability realization that is currently being used by the transmitter and the constellation distribution for that prior realization. As this generic formulation involves optimization over a space of joint PDFs, an alternative optimization problem, the optimal solution of which can be expressed as a randomization among at most three mass points, is derived, and it is proved that the original and the alternative problems share the same optimal value. Next, three special cases of the original formulation are investigated. First, the optimal prior distribution for a given constellation is derived. Second, the optimal pair of fixed priors and signal levels is considered, and third, a binary communication scenario with scalar observations under additive zero-mean Gaussian noise is investigated. Finally, numerical results are provided for the general formulation and the special cases. The main contributions in this paper can be summarized as follows:

- For the first time in the literature, the optimal signaling problem is proposed by jointly optimizing the signal constellation and the prior probabilities for transmitted symbols in the presence of average bit rate and power constraints.
- It is shown that the optimal performance is achieved by a randomization among at most three signal constellations with the corresponding associated deterministic prior probability vectors.
- A closed form expression of the optimal deterministic prior probability distribution for a given constellation is derived.
- The optimal solution for the special case of binary communications over an AWGN channel with scalar observations is obtained under certain practical assumptions.

The rest of the paper is organized as follows: The optimal signaling problem is formulated and form of the solution is provided in Section II. Special cases of the general formulation are discussed in Section III. Numerical results are presented in Section IV and concluding remarks are given in Section V.

II. FORMULATION AND OPTIMAL SIGNALING

Consider an M-ary communication system with n dimensional observations collected at the receiver over an arbitrary additive noise channel. The discrete-time baseband equivalent signal after downconversion, matched filtering, and sampling at the symbol rate can be represented as

$$\mathbf{Y} = \mathbf{S}_i + \mathbf{N}, \quad i \in \{0, 1, \ldots, M-1\} \tag{1}$$

where $\mathbf{S}_i$ is the transmitted signal vector for ith constellation symbol and $\mathbf{N}$ denotes the noise vector that is assumed to be independent of $\mathbf{S}_i$. Prior probabilities of the symbols are denoted by $\mathbf{\Pi} := [\Pi_0, \ \Pi_1, \ldots, \ \Pi_{M-1}]$, which belongs to the standard $(M-1)-$simplex denoted with $\Delta^{M-1} = \{\boldsymbol{\pi} : \sum_{i=0}^{M-1} \pi_i = 1 \text{ and } \pi_i \geq 0 \text{ for all } i\}$. We recall that the standard simplex is a compact and convex set. Our goal is

to obtain the optimal distribution for the prior probabilities and the transmitted symbols that maximize the probability of correct decision at the receiver subject to constraints on the average transmit power and the average bit rate. To this end, the prior probability vector $\mathbf{\Pi}$ and the transmitted symbols $\mathbf{S}_i$'s are assumed to be random with a joint distribution denoted by $p_{\mathbf{\Pi},\mathbf{S}}(\boldsymbol{\pi},\mathbf{s})$ where $\mathbf{S} := [\mathbf{S}_0, \ \mathbf{S}_1, \ \ldots, \ \mathbf{S}_{M-1}] \in \mathbb{R}^{Mn}$ represents the signal constellation. The average transmit power constraint and the average bit rate per symbol constraint are given by

$$\mathbb{E}\left\{\sum_{i=0}^{M-1} \Pi_i \big\|\mathbf{S}_i\big\|_2^2\right\} \leq A, \qquad (2)$$

and

$$\mathbb{E}\left\{-\sum_{i=0}^{M-1} \Pi_i \log \Pi_i\right\} \geq R, \qquad (3)$$

respectively. In (2) and (3), the expectations are taken with respect to the joint PDF $p_{\mathbf{\Pi},\mathbf{S}}(\boldsymbol{\pi},\mathbf{s})$. It is noted that for a given prior probability vector $\boldsymbol{\pi}$ and a signal constellation $\mathbf{s}$, the optimal detector at the receiver corresponds to the MAP decision rule [2, Theorem 2.7.3]. More specifically, for a given observation $\mathbf{y}$, the MAP decision rule selects symbol k such that $k = \arg\max_{i \in \{0,1,\ldots,M-1\}} \pi_i p_i(\mathbf{y})$, where $p_i(\mathbf{y})$ denotes the conditional PDF of the observation when the ith symbol is transmitted. The transmitter and the receiver are assumed to be in coordination so that the receiver knows which prior probability vector is currently being used by the transmitter. Accordingly, the average probability of correct decision can be expressed as

$$\mathrm{P_c} := \mathbb{E}\left\{\int_{\mathbb{R}^n} \max_{i \in \{0,1,\ldots,M-1\}} \left\{\Pi_i \ \mathbb{E}\left\{p_\mathbf{N}(\mathbf{y}-\mathbf{S}_i) \,|\, \mathbf{\Pi}\right\}\right\} d\mathbf{y}\right\}, \qquad (4)$$

where the outer expectation is taken with respect to the marginal PDF of $\mathbf{\Pi}$, that is, $p_{\mathbf{\Pi}}(\boldsymbol{\pi})$, and the inner expectation is taken with respect to the conditional PDF of $\mathbf{S}$ given $\mathbf{\Pi}$, i.e., $p_{\mathbf{S}|\mathbf{\Pi}}(\mathbf{s}|\boldsymbol{\pi})$. Then, the following optimization problem is proposed:

$$\max_{p_{\mathbf{\Pi},\mathbf{s}}} \ \mathbb{E}\left\{\int_{\mathbb{R}^n} \max_{i \in \{0,1,\ldots,M-1\}} \left\{\Pi_i \mathbb{E}\left\{p_\mathbf{N}(\mathbf{y}-\mathbf{S}_i) \,|\, \mathbf{\Pi}\right\}\right\} d\mathbf{y}\right\}$$

$$\text{subject to} \ \ \mathbb{E}\left\{\sum_{i=0}^{M-1} \Pi_i \big\|\mathbf{S}_i\big\|_2^2\right\} \leq A$$

$$\mathbb{E}\left\{-\sum_{i=0}^{M-1} \Pi_i \log \Pi_i\right\} \geq R \qquad \text{(P1)}$$

where the optimization is over the joint PDF $p_{\mathbf{\Pi},\mathbf{S}}(\boldsymbol{\pi},\mathbf{s})$. Note that in (P1), focusing on the objective function, if $\mathbf{\Pi}$ is taken to be a fixed deterministic probability vector, then the problem reduces to the optimal stochastic signaling problem with the corresponding MAP detector employed at the receiver [18]. On the other hand, if the constellation $\mathbf{S}$ is fixed, then the problem simplifies to finding the optimal randomization over multiple MAP detectors [16].

As (P1) involves optimization in the space of joint PDFs, it

is in general difficult to solve. In the following, an upper bound on the objective function of (P1) is obtained by interchanging maximum and expectation operations, and the form of the solution is characterized for the resulting problem. Then, it is shown that the original problem has the same solution as that of the one based on the upper bound. To this aim, consider the following objective function:

$$\mathrm{P_c} := \mathbb{E}\left\{\int_{\mathbb{R}^n} \max_{i \in \{0,1,\ldots,M-1\}} \left\{\pi_i \ p_\mathbf{N}(\mathbf{y}-\mathbf{S}_i)\right\} d\mathbf{y}\right\}, \qquad (5)$$

where the expectation is taken with respect to the joint PDF $p_{\mathbf{\Pi},\mathbf{S}}(\boldsymbol{\pi},\mathbf{s})$. Then, based on (P1) and (5), an alternative optimization problem is formulated as

$$\max_{p_{\mathbf{\Pi},\mathbf{s}}} \ \mathbb{E}\left\{\int_{\mathbb{R}^n} \max_{i \in \{0,1,\ldots,M-1\}} \left\{\Pi_i \ p_\mathbf{N}(\mathbf{y}-\mathbf{S}_i)\right\} d\mathbf{y}\right\}$$

$$\text{subject to} \ \ \mathbb{E}\left\{\sum_{i=0}^{M-1} \Pi_i \big\|\mathbf{S}_i\big\|_2^2\right\} \leq A$$

$$\mathbb{E}\left\{-\sum_{i=0}^{M-1} \Pi_i \log \Pi_i\right\} \geq R \qquad \text{(P2)}$$

Remark 1: The formulation in (P2) corresponds to the scenario in which the receiver and the transmitter are fully coordinated about the transmission policy. More specifically, the receiver is informed of the constellation and the corresponding prior probability vector employed at the transmitter at any given instant. Hence, the optimal decision rule can be implemented at the receiver. For example, in a slotted communication scenario, this can be realized by assigning each slot with a designated prior distribution and a signal constellation, and allocating the number of slots corresponding to that realization in proportion to its weight in the joint PDF.

The optimization problem in (P2) can be expressed in a more compact form. To this end, define the random vector $\mathbf{X}$ as follows:

$$\mathbf{X} := [\mathbf{\Pi}, \ \mathbf{S}] = [\Pi_0, \ \Pi_1, \ldots, \Pi_{M-1}, \ \mathbf{S}_0, \ \mathbf{S}_1, \ldots, \mathbf{S}_{M-1}] \qquad (6)$$

where $\mathbf{X} \in \Delta^{M-1} \times \mathbb{R}^{Mn}$. Then, (P2) can equivalently be expressed as

$$\max_{p_\mathbf{X}} \ \mathbb{E}\left\{F(\mathbf{X})\right\}$$

$$\text{subject to} \ \mathbb{E}\left\{G(\mathbf{X})\right\} \leq A$$

$$\mathbb{E}\left\{H(\mathbf{X})\right\} \geq R \qquad (7)$$

with

$$F(\mathbf{X}) := \int_{\mathbb{R}^n} \max_{i \in \{0,1,\ldots,M-1\}} \left\{\Pi_i \ p_\mathbf{N}(\mathbf{y}-\mathbf{S}_i)\right\} d\mathbf{y},$$

$$G(\mathbf{X}) := \sum_{i=0}^{M-1} \Pi_i \big\|\mathbf{S}_i\big\|_2^2,$$

$$H(\mathbf{X}) := -\sum_{i=0}^{M-1} \Pi_j \log \Pi_i,$$

where the expectations are taken with respect to the joint PDF of the constellation points and prior probabilities denoted by $p_\mathbf{X}(\boldsymbol{x})$. Note that there are also implicit constraints

in (7), that is, $p_{\mathbf{X}}(\mathbf{x}) \geq 0 \ \forall \mathbf{x} \in \Delta^{M-1} \times \mathbb{R}^{Mn}$ and $\int_{\Delta^{M-1} \times \mathbb{R}^{Mn}} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1$ must be satisfied. In (7), $F(\mathbf{x})$ with $\mathbf{x} = [\boldsymbol{\pi}, \ \mathbf{s}]$ can be viewed as the probability of correct decision when a fixed deterministic constellation $\mathbf{s}$ is used for the transmission of M symbols whose prior probabilities are specified by $\boldsymbol{\pi}$ and the corresponding MAP detector is employed at the receiver.

Optimization problems in the form of (7) have been studied in the literature [14], [16]-[20]. If $F(\mathbf{x})$ is continuous and the components of $\mathbf{x}$ belong to finite closed intervals, then the optimal solution of (7) can be expressed as a randomization among at most three points, which follows from Carethéodory's theorem [13], [28]. Therefore, instead of searching over the space of all PDFs, we can restrict the search for the optimal solution to a family of PDFs in the form $p_{\mathbf{X}}^{\mathrm{opt}}(\mathbf{x}) = \sum_{j=1}^{3} \lambda_j \delta(\mathbf{x} - \mathbf{x}_j)$ where δ denotes the Dirac delta function, $\sum_{j=1}^{3} \lambda_j = 1$ and $\lambda_j \geq 0 \ \forall j$. Based on this result, the optimization problem in (7) can be simplified to

$$\max_{\{\lambda_1, \lambda_2, \lambda_3, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}} \sum_{j=1}^{3} \lambda_j F(\mathbf{x}_j)$$
$$\text{subject to } \sum_{j=1}^{3} \lambda_j G(\mathbf{x}_j) \leq A, \quad \sum_{j=1}^{3} \lambda_j H(\mathbf{x}_j) \geq R,$$
$$\sum_{j=1}^{3} \lambda_j = 1, \quad \lambda_1, \lambda_2, \lambda_3 \geq 0 \tag{8}$$

where $F(.)$, $G(.)$, and $H(.)$ are as defined before, $\mathbf{x}_j = [\pi_{j,0}, \ \pi_{j,1}, \dots \ \pi_{j,M-1}, \ \mathbf{s}_{j,0}, \ \mathbf{s}_{j,1}, \dots, \mathbf{s}_{j,M-1}]$ and $s_{j,i}$ is the ith symbol in the jth signal constellation. Next, the following proposition is presented.

Proposition 1: *Given the same average power constraint A, bit rate constraint R, and the noise PDF $p_{\mathbf{N}}(\cdot)$, the optimization problems in* (P1) *and* (P2) *have the same optimal value.*

Proof: Denote the optimal values of the optimization problems in (P1) and (P2) as P_c^* and the $P_c^\dagger$, respectively. We first establish $P_c^* \leq P_c^\dagger$. For any given joint distribution $p_{\boldsymbol{\Pi}, \mathbf{s}}$,

$$\mathbb{E}\left\{ \int_{\mathbb{R}^n} \max_{i \in \{0,1,\dots,M-1\}} \left\{ \Pi_i \, \mathbb{E}\left\{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{S}_i) \,|\, \boldsymbol{\Pi} \right\} \right\} dy \right\}$$
$$\leq \mathbb{E}\left\{ \mathbb{E}\left\{ \int_{\mathbb{R}^n} \max_{i \in \{0,1,\dots,M-1\}} \left\{ \Pi_i \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{S}_i) \right\} dy \,\Big|\, \boldsymbol{\Pi} \right\} \right\}$$
$$= \mathbb{E}\left\{ \int_{\mathbb{R}^n} \max_{i \in \{0,1,\dots,M-1\}} \left\{ \Pi_i \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{S}_i) \right\} dy \right\} \tag{9}$$

where the inequality follows by interchanging the order of the inner maximization and expectation operators and the equality is due to the law of total expectation. Hence, under the same feasible set of joint PDFs, the optimal values of the objective functions in problems (P1) and (P2) satisfy $P_c^* \leq P_c^\dagger$. Next, we show that $P_c^* \geq P_c^\dagger$. Consider the joint PDF for the form of the optimal solution of (P2), i.e., $p_{\boldsymbol{\Pi}, \mathbf{s}}(\boldsymbol{\pi}, \mathbf{s}) = \sum_{j=1}^{3} \lambda_j p_{\boldsymbol{\Pi}, \mathbf{s}}^{(j)}(\boldsymbol{\pi}, \mathbf{s})$ with $p_{\boldsymbol{\Pi}, \mathbf{s}}^{(j)}(\boldsymbol{\pi}, \mathbf{s}) = p_{\boldsymbol{\Pi}}^{(j)}(\boldsymbol{\pi}) p_{\mathbf{s}|\boldsymbol{\Pi}}^{(j)}(\mathbf{s}|\boldsymbol{\pi})$, where $p_{\boldsymbol{\Pi}}^{(j)}(\boldsymbol{\pi}) = \delta(\boldsymbol{\pi} - \boldsymbol{\pi}_j)$, $\boldsymbol{\pi}_j = [\pi_{j,0}, \ \pi_{j,1}, \dots, \ \pi_{j,M-1}]$, $p_{\mathbf{s}|\boldsymbol{\Pi}}^{(j)}(\mathbf{s}|\boldsymbol{\pi}) = p_{\mathbf{s}|\boldsymbol{\Pi}}(\mathbf{s}|\boldsymbol{\pi}_j) = \delta(\mathbf{s} -$

$\mathbf{s}_j)$, and $\mathbf{s}_j = [\mathbf{s}_{j,0}, \ \mathbf{s}_{j,1}, \dots, \ \mathbf{s}_{j,M-1}]$. When this PDF is employed, (P1) reduces to (P2). However, since this is just a special case for the solution of (P1), one obtains $P_c^* \geq P_c^\dagger$. Therefore, it is concluded that $P_c^* = P_c^\dagger$. $\blacksquare$

Remark 2: It should be noted that employing a signaling scheme with nonuniform priors results in variable-rate data transmission since the number of bits transmitted during a signaling interval is a random variable. Hence, it is susceptible to buffer over- or underflow for a fixed-rate source as well as synchronization loss due to channel errors causing insertion and deletion of bits in the decoded data. In practice, near optimal nonuniform signaling schemes can be designed by parsing a binary data stream into the codewords of the variable-length prefix code designed using the Huffman algorithm and then mapping them onto the points of the given constellation.

Remark 3: By following the transmission protocol explained in Remark 1, the randomization idea can be implemented based on $p_{\mathbf{X}}^{\mathrm{opt}}(\mathbf{x})$. It is interesting to note that if the transmitted symbols are observed over a long duration, it would be as if the transmission is performed over a larger deterministic constellation $\hat{\mathbf{x}} = [\lambda_1 \pi_{1,0}, \dots, \ \lambda_1 \pi_{1,M-1}, \dots, \lambda_3 \pi_{1,0}, \ \dots, \ \lambda_3 \pi_{3,M-1}, \mathbf{s}_{1,0}, \dots, \mathbf{s}_{1,M-1}, \dots, \mathbf{s}_{3,0}, \dots, \mathbf{s}_{3,M-1}]$. By introducing certain protocols between the transmitter and the receiver to implement the M-ary communication system based on $\hat{\mathbf{x}}$ (while satisfying the average bit rate (defined for the M-ary system) and power constraints), the optimization problem can be regarded as a search of the optimal deterministic vector $\hat{\mathbf{x}}$. However, both the randomization idea formulated in this paper or this alternative approach are actually equivalent and would yield the same system performance.

III. Special Cases

A. Optimal Deterministic Prior Distribution for Given Constellation

In this section, we provide a closed-form solution for the optimal deterministic prior distribution for a given signal constellation. Consider a communication system in which the transmitter emits a sequence of symbols drawn independently from a fixed constellation $\Omega = \{\mathbf{s}_0, \dots, \mathbf{s}_{M-1}\} \subset \mathbb{R}^{Mn}$. The (deterministic) prior probability vector of the signals is denoted by $\boldsymbol{\pi}$. Under these assumptions, the optimization problem can be formulated as (cf. (7))

$$\max_{\boldsymbol{\pi} \in \Delta^{M-1}} F(\boldsymbol{\pi})$$
$$\text{subject to } H(\boldsymbol{\pi}) \geq R$$
$$G(\boldsymbol{\pi}) \leq A \tag{10}$$

where $F(\boldsymbol{\pi}) = \int_{\mathbb{R}_n} \max_{i \in \{0,1,\dots,M-1\}} \{\pi_i \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_i)\} dy$, $G(\boldsymbol{\pi}) = \sum_{i=0}^{M-1} \pi_i \|\mathbf{s}_i\|^2$, and $H(\boldsymbol{\pi}) = -\sum_{i=0}^{M-1} \pi_i \log_2(\pi_i)$. We recall that $H(\boldsymbol{\pi})$ is a concave function of $\boldsymbol{\pi}$ and attains a maximum value of $\log_2 M$ in the case of uniform signaling, i.e., when $\pi_i = 1/M$ for all $i = 0, \dots, M-1$ [29, Theorem 2.7.3]. On the other hand, $G(\boldsymbol{\pi})$ is a linear function of $\boldsymbol{\pi}$ and $F(\boldsymbol{\pi})$ is a convex function of $\boldsymbol{\pi}$, which follows from the fact that the minimum Bayes error is a concave function

of $\boldsymbol{\pi}$ over the standard simplex [2, Section II.C]. In (10), it is required that the constellation Ω must be able to support the average power A, i.e., $A \geq A_{\min}$, where $A_{\min}$ is the power of a minimum-power point in Ω. Additionally, $0 \leq R \leq \tilde{R}(A)$ is needed for feasibility, where $\tilde{R}(A)$ is the maximum average bit rate that can be attained under the average symbol power constraint A [22].

1) Proposed Solution: The proposed approach for solving the optimization problem in (10) is to first characterize the form of the solution for an arbitrary detector at the receiver and then to apply the optimal MAP decision rule. To that aim, we consider a generic detector at the receiver specified by the decision functions $\boldsymbol{\delta} := (\delta_0, \ldots, \delta_{M-1})$. Upon the reception of an observation $\boldsymbol{y}$, the receiver decides in favor of the hypothesis that $\boldsymbol{s}_i$ is transmitted with probability $\delta_i(\boldsymbol{y})$, where $\delta_i(\boldsymbol{y}) \geq 0$ and $\sum_{i=0}^{M-1} \delta_i(\boldsymbol{y}) = 1$ for all $\boldsymbol{y} \in \mathbb{R}^n$. For a given detector $\boldsymbol{\delta}$ and signaling probabilities $\boldsymbol{\pi}$, the average correct decision probability is expressed as $P_c(\boldsymbol{\pi}, \boldsymbol{\delta}) = \sum_{i=0}^{M-1} \pi_i P_{c,i}(\delta_i)$, where $P_{c,i}(\delta_i)$ denotes the average probability of correct decision given that $\boldsymbol{s}_i$ is transmitted, i.e.,

$$P_{c,i}(\delta_i) = \mathbb{E}_i \{\delta_i(\boldsymbol{Y})\} = \int_{\mathbb{R}^n} \delta_i(\boldsymbol{y}) p_i(\boldsymbol{y}) d\boldsymbol{y}$$
$$= \int_{\mathbb{R}^n} \delta_i(\boldsymbol{y}) p_{\mathbf{N}}(\boldsymbol{y} - \mathbf{s}_i) d\boldsymbol{y} \qquad (11)$$

Next, we present the following lemma.

Lemma 1: For a given detector specified by the decision functions $\{\delta_i\}_{i=1}^M$, the following signaling distribution

$$\pi_i^* = \exp\left(-\lambda_1 ||\boldsymbol{s}_i||^2 + \lambda_2 P_{c,i}(\delta_i)\right) / Z(\lambda_1, \lambda_2), \qquad (12)$$

for $i = 0, \ldots, M-1$, where $\lambda_1, \lambda_2 \geq 0$ and $Z(\lambda_1, \lambda_2) = \sum_{i=0}^{M-1} \exp\left(-\lambda_1 ||\boldsymbol{s}_i||^2 + \lambda_2 P_{c,i}(\delta_i)\right)$, maximizes the average probability of correct decision under constraints on average bit rate and average symbol power.

Proof: For a given detector, the problem in (10) takes the following form:

$$\max_{\boldsymbol{\pi}} \sum_{i=0}^{M-1} \pi_i P_{c,i}(\delta_i)$$

$$\text{subject to} \quad -\sum_{i=0}^{M-1} \pi_i \log_2(\pi_i) \geq R \qquad (13a)$$

$$\sum_{i=0}^{M-1} \pi_i ||\boldsymbol{s}_i||^2 \leq A, \qquad (13b)$$

$$\sum_{i=0}^{M-1} \pi_i = 1, \ \pi_i \geq 0, \ i = 0, \ldots, M-1 \qquad (13c)$$

Notice that Slater's conditions hold for the optimization problem in (13). More explicitly, the optimization in (13) is convex and for $R < \log_2 M$, the non-affine inequality constraint in (13a) is strictly satisfied with $\pi_i = 1/M$, $i = 0, \ldots, M-1$. Hence, strong duality holds and Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient [30]. The Lagrangian

function corresponding to the optimization problem in (13) is

$$L(\boldsymbol{\pi}; \gamma_1, \gamma_2, \nu) = \sum_{i=0}^{M-1} \pi_i P_{c,i}(\delta_i) - \gamma_1 \left(\sum_{i=0}^{M-1} \pi_i \log_2(\pi_i) + R \right)$$
$$+ \gamma_2 \left(A - \sum_{i=0}^{M-1} \pi_i ||\boldsymbol{s}_i||^2 \right) + \nu \left(\sum_{i=0}^{M-1} \pi_i - 1 \right). \qquad (14)$$

Taking the derivative with respect to π_i and equating to zero yields

$$\pi_i^* = 2^{-\log_2 e + \left(P_{c,i}(\delta_i) - \gamma_2 ||\boldsymbol{s}_i||^2 + \nu\right)/\gamma_1}. \qquad (15)$$

Applying the condition $\sum_{i=0}^{M-1} \pi_i = 1$ and reparameterizing with $\lambda_1 = (\gamma_2/\gamma_1) \ln 2$ and $\lambda_2 = (\ln 2)/\gamma_1$, we get

$$\pi_i^* = \exp\left(-\lambda_1 ||\boldsymbol{s}_i||^2 + \lambda_2 P_{c,i}(\delta_i)\right) / Z(\lambda_1, \lambda_2) \qquad (16)$$

where $Z(\lambda_1, \lambda_2) = \sum_{i=0}^{M-1} \exp\left(-\lambda_1 ||\boldsymbol{s}_i||^2 + \lambda_2 P_{c,i}(\delta_i)\right)$ and $\lambda_1, \lambda_2 \geq 0$ follows from the dual feasibility condition, i.e., $\gamma_1, \gamma_2 \geq 0$. ∎

The parameters λ_1 and λ_2 govern the trade-off among the average probability of correct decision, the average bit rate, and the average symbol power. For fixed λ_2, as λ_1 is increased, the inner constellation points (i..e, those with low power) are selected more frequently than the outer constellation points (i.e., those with high power). On the other hand, for fixed λ_1, as λ_2 is increased, constellation points yielding lower symbol error probability are selected more frequently than those with higher error rates.[1] In addition, constellation points that have the same power and the same error probability are selected equally likely. Lastly, we note that the signaling distribution that maximizes the average bit rate under the average symbol power constraint (equivalently, minimizes the average power for a fixed bit rate) can be obtained by substituting $\lambda_2 = 0$ and solving for λ_1 from the power constraint [22]. In light of the lemma, the following proposition characterizes the optimal signaling distribution that solves the optimization in (10).

Proposition 2: For any given R as the upper bound on the average symbol power that is supported by a given constellation Ω and $R \leq \tilde{R}(A)$ as the lower bound on the average bit rate, where $\tilde{R}(A)$ is the maximum average bit rate that can be attained under an average symbol power constraint A, the solution $\boldsymbol{\pi}^* = (\pi_0^*, \ldots, \pi_{M-1}^*)$ to (10) satisfies the following equation (i.e., a fixed point):

$$\pi_i^* = \frac{\exp\left(-\lambda_1^* ||\boldsymbol{s}_i||^2 + \lambda_2^* P_{c,i}(\delta_i^*)\right)}{\sum_{j=1}^M \exp\left(-\lambda_1^* ||\boldsymbol{s}_j||^2 + \lambda_2^* P_{c,j}(\delta_j^*)\right)} \qquad (17)$$

for $i = 0, \ldots, M-1$, where $\boldsymbol{\delta}^* = \{\delta_i^*\}_{i=0}^{M-1}$ is the MAP detector corresponding to the optimal signaling distribution $\boldsymbol{\pi}^*$, i.e.,

$$\delta_i^*(\boldsymbol{y}) = 1, \quad if \ i = \underset{k \in \{0, \ldots, M-1\}}{\arg\max} \pi_k^* p_k(\boldsymbol{y}) \qquad (18)$$

and $\delta_i^*(\boldsymbol{y}) = 0$ otherwise, for $i = 0, \ldots, M-1$ and every

---

[1]In general, a lower symbol error probability can be achieved by selecting a fewer number of constellation points that are farther apart from each other (e.g., at the vertices of the constellation). In the limit as $\lambda_2 \to \infty$, this would result in degenerate signaling (i.e., $\pi_i = 1$ for some $i \in \{1, \ldots, M\}$ yielding zero bit rate.)

$\boldsymbol{y} \in \mathbb{R}^n$. The optimal parameters λ_1^* and λ_2^* are obtained as follows:

Case 1: Let $\lambda_1^* = 0$ and $\lambda_2^* \geq 0$ be a solution to

$$-\sum_{i=0}^{M-1} \pi_i(\lambda_2) \log_2(\pi_i(\lambda_2)) = R \qquad (19)$$

where $\boldsymbol{\pi}(\lambda_2) = (\pi_0(\lambda_2), \ldots, \pi_{M-1}(\lambda_2))$ satisfies $\pi_i(\lambda_2) = \exp(\lambda_2 P_{c,i}(\delta_i)) / \left(\sum_{j=1}^{M} \exp(\lambda_2 P_{c,j}(\delta_j)) \right)$, $i = 0, \ldots, M-1$ and $\boldsymbol{\delta} = \{\delta_i\}_{i=0}^{M-1}$ is the MAP detector corresponding to $\boldsymbol{\pi}(\lambda_2)$. Then, $\{\boldsymbol{\pi}^*(\lambda_2^*), \lambda_2^*\}$ together with $\lambda_1^* = 0$ is optimal if the constraint on the average symbol power is satisfied, i.e.,

$$\sum_{i=0}^{M-1} \pi_i^*(\lambda_2^*) \|\boldsymbol{s}_i\|^2 \leq A, \qquad (20)$$

else if (20) fails, go to Case 2.

Case 2: Let $\lambda_1^* > 0$ and $\lambda_2^* \geq 0$ be a solution to

$$-\sum_{i=0}^{M-1} \pi_i(\lambda_1, \lambda_2) \log_2(\pi_i(\lambda_1, \lambda_2)) = R,$$

$$\sum_{i=0}^{M-1} \pi_i(\lambda_1, \lambda_2) \|\boldsymbol{s}_i\|^2 = A \qquad (21)$$

where $\boldsymbol{\pi}(\lambda_1, \lambda_2) = (\pi_0(\lambda_1, \lambda_2), \ldots, \pi_{M-1}(\lambda_1, \lambda_2))$ satisfies

$$\pi_i(\lambda_1, \lambda_2) = \frac{\exp\left(-\lambda_1 \|\boldsymbol{s}_i\|^2 + \lambda_2 P_{c,i}(\delta_i)\right)}{\sum_{j=1}^{M} \exp\left(-\lambda_1 \|\boldsymbol{s}_j\|^2 + \lambda_2 P_{c,j}(\delta_j)\right)} \qquad (22)$$

and $\boldsymbol{\delta} = \{\delta_i\}_{i=0}^{M-1}$ is the MAP detector corresponding to $\boldsymbol{\pi}(\lambda_1, \lambda_2)$. Then, $\{\boldsymbol{\pi}^*(\lambda_1^*, \lambda_2^*), \lambda_1^*, \lambda_2^*\}$ is optimal.

Proof: Please see Appendix A.

Since the optimal signaling distribution $\boldsymbol{\pi}(\lambda_1, \lambda_2)$ is a continuous function of λ_1 and λ_2, an iterative bisection search algorithm can be employed to solve for the values of λ_1 and λ_2 that satisfy the equality constraints in (19) and (21).

B. Joint Design of Optimal Deterministic Priors and Constellation Points

In this section, we formulate the problem of jointly designing optimal *deterministic* signal constellation and the corresponding prior probabilities of the constellation symbols. Namely, instead of searching for the optimal PDF as specified by the general problem in (8), we try to find the single point $\boldsymbol{x} = [\boldsymbol{\pi}, \boldsymbol{s}_0, \ldots, \boldsymbol{s}_{M-1}] \in \Delta^{M-1} \times \mathbb{R}^{Mn}$ that maximizes the average probability of correct decision under average transmission power and bit rate per symbol constraints. Therefore, the optimization problem can be formulated as (cf. (7))

$$\max_{\boldsymbol{x} \in \Delta^{M-1} \times \mathbb{R}^{Mn}} F(\boldsymbol{x})$$
$$\text{subject to } H(\boldsymbol{x}) \geq R$$
$$G(\boldsymbol{x}) \leq A \qquad (23)$$

where $F(\boldsymbol{x}) = \int_{\mathbb{R}_n} \max_{i \in \{0,1,\ldots,M-1\}} \{\pi_i \, p_{\mathbf{N}}(\boldsymbol{y} - \boldsymbol{s}_i)\} d\boldsymbol{y}$, $G(\boldsymbol{x}) = \sum_{i=0}^{M-1} \pi_i \|\boldsymbol{s}_i\|^2$, and $H(\boldsymbol{x}) = -\sum_{i=0}^{M-1} \pi_i \log_2(\pi_i)$. Notice that if the signal constellation $\boldsymbol{s} = \{\boldsymbol{s}_0, \ldots, \boldsymbol{s}_{M-1}\} \subset \mathbb{R}^{Mn}$ is fixed in $\boldsymbol{x}$, then the problem in (23) reduces to that in

(10). As the solution is known for the prior distribution for a given $\boldsymbol{s}$, average power constraint A, and bit rate constraint R based on Proposition 2, one can actually perform the optimization over the signal constellation $\boldsymbol{s}$ only. Let $\boldsymbol{\pi}^*(\boldsymbol{s})$ denote the optimal prior distribution for the signal constellation $\boldsymbol{s}$, which can be obtained according to Case 1 or Case 2 in Proposition 2. Then, (23) becomes

$$\max_{\boldsymbol{s} \in \mathbb{R}^{Mn}} \int_{\mathbb{R}_n} \max_{i \in \{0,1,\ldots,M-1\}} \{\pi_i^*(\boldsymbol{s}) \, p_{\mathbf{N}}(\boldsymbol{y} - \boldsymbol{s}_i)\} d\boldsymbol{y}. \qquad (24)$$

Note that for some $\boldsymbol{s} \in \mathbb{R}^{Mn}$, the reduced problem of optimal prior distribution may not be feasible for given A and R; hence, $\boldsymbol{\pi}^*(\boldsymbol{s})$ may not exist. In that case, one can simply set the objective function in (24) to take the value $-\infty$.

Remark 4: Let $\boldsymbol{x}^{\text{opt}}$ denote the optimal solution to (23). Then, $H(\boldsymbol{x}^{\text{opt}}) = R$. This immediately follows from the form of the solution to $\boldsymbol{\pi}^*$ given in Proposition 2.

C. Binary Communication over AWGN Channel

In this section, we investigate the special case of a binary communication system with scalar observations, corrupted by a zero-mean Gaussian noise with variance σ^2. In this case, we get $\mathbf{X} = [\Pi_0, \Pi_1, S_0, S_1]$, where $\Pi_0 = 1 - \Pi_1$. It is assumed that for any given realization $\mathbf{X} = \mathbf{x}_i$, $G(\mathbf{x}_i) \leq A$ holds; that is, an individual power constraint is imposed for each pair of constellation set and the corresponding prior probability vector.

In the absence of the bit rate constraint, it is well-known that for given prior probabilities (π_0, π_1), the optimal constellation symbols that minimize the probability of error, in the presence of the MAP detector and average power constraint A, are $S_0 = -\sqrt{A}/\alpha$ and $S_1 = \alpha\sqrt{A}$ with $\alpha = \sqrt{\pi_0/\pi_1}$ when the noise distribution is Gaussian [19]. To this end, when there exist average power and bit rate constraints on the signal, the optimization over the distribution of $\mathbf{X}$, can be reduced to an optimization over the distribution of Π_1, since the optimal signal constellation is well-defined for any given prior realization. This implies that the average power constraint can be omitted, as it always holds with equality. Therefore, let $p_{\Pi_1}(\pi_1)$ denote the PDF of prior Π_1 corresponding to symbol S_1. Then, the problem can be expressed in terms of minimization of the probability of error as follows:

$$\min_{p_{\Pi_1}} \mathbb{E}(f(\Pi_1)) \quad \text{subject to } \mathbb{E}(h(\Pi_1)) \geq R, \qquad (25)$$

with $f(\pi_1) \triangleq \int_{-\infty}^{\infty} \min\{\pi_1 \, p_N(y - \alpha\sqrt{A}), (1 - \pi_1) \, p_N(y + \sqrt{A}/\alpha)\} dy$ and $h(\pi_1) \triangleq -\pi_1 \log \pi_1 - (1 - \pi_1) \log(1 - \pi_1)$, where the expectations are taken with respect to $p_{\Pi_1}(\pi_1)$ and $p_N(y) = (1/\sqrt{2\pi\sigma^2}) \, e^{-y^2/2\sigma^2}$. For the Gaussian noise, the optimal MAP detector is a single threshold detector. Then, $f(\pi_1)$ can be expressed as

$$f(\pi_1) = \pi_1 \int_{-\infty}^{\tau(\pi_1)} p_N(y - \alpha\sqrt{A}) \, dy$$
$$+ (1 - \pi_1) \int_{\tau(\pi_1)}^{\infty} p_N(y + \sqrt{A}/\alpha) \, dy \qquad (26)$$

where $\tau(\pi_1) = 0.5\sqrt{A}(\alpha - 1/\alpha) + \frac{2\sigma^2 \ln(\alpha)}{\sqrt{A}(\alpha+1/\alpha)}$ with $\alpha \triangleq \sqrt{(1-\pi_1)/\pi_1}$ [2]. Note that both $f(\pi_1)$ and $h(\pi_1)$ are symmetric around $\pi_1 = 0.5$; thus, we can restrict the values of prior π_1 to the interval $[0, 0.5]$. In this region, $h(\pi_1)$ is a monotone concave function of π_1; hence, its inverse function exists. Let h^{-1} denote the inverse entropy function with $h^{-1} : [0, 1] \to [0, 0.5]$ and $h^{-1}(r) = \pi_1$ when $h(\pi_1) = r$ for $r \in [0, 1]$ and $\pi_1 \in [0, 0.5]$. Note that $f(\pi_1)$ can be rewritten as

$$
\begin{aligned}
&f(\pi_1) \\
&= \pi_1 Q\left(\frac{\alpha\sqrt{A} - \tau(\pi_1)}{\sigma}\right) + (1 - \pi_1) Q\left(\frac{\sqrt{A}/\alpha + \tau(\pi_1)}{\sigma}\right) \\
&= \pi_1 Q\left(\gamma \frac{\alpha^2 + 1}{2\alpha} - \frac{2\ln\alpha}{\gamma(\alpha + 1/\alpha)}\right) \\
&\quad + (1 - \pi_1) Q\left(\gamma \frac{\alpha^2 + 1}{2\alpha} + \frac{2\ln\alpha}{\gamma(\alpha + 1/\alpha)}\right)
\end{aligned}
\tag{27}
$$

where $\gamma \triangleq \sqrt{A}/\sigma$. Note that f depends only on γ and π_1. Based on the preceding definitions, the following results are presented.

Property 1: $g(r)$ *is a strictly convex function on* $[0, 1]$ *for* $\gamma > \gamma_{th} \approx 0.166$.

Derivation: Please see Appendix B.

Lemma 2: Let $g(r) = f \circ h^{-1}(r)$. Then, $g(r)$ is monotone increasing on $[0, 1]$ for $\gamma > 0$.

Proof: Please see Appendix C.

Property 2: *Under individual power constraint A on each pair of signal constellations and the corresponding prior probability vector, for a given average bit rate constraint R and $\gamma > \gamma_{th} \approx 0.166$, the optimal prior probability distribution for a binary communication system with an additive Gaussian noise channel does not involve randomization and can be specified as $p_{\Pi_1}^{\mathrm{opt}}(\pi_1) = \delta(\pi_1 - h^{-1}(R))$. The corresponding optimal constellation can be specified as $(S_0, S_1) = (-\sqrt{A}/\alpha, \alpha\sqrt{A})$ with $\alpha \triangleq \sqrt{\pi_0/\pi_1}$ and $\pi_1 = h^{-1}(R)$.*

Derivation: It is first noted that $g(r)$ is monotone increasing and strictly convex when $\gamma > \gamma_{th} \approx 0.166$. Under the constraint that $h(\pi_1) \geq R$, we have $h^{-1}(R) = \arg\max_{\pi_1} f(\pi_1)$ due to monotonicity. Assume that there exists a PDF $p_{\Pi_1}^{\mathrm{opt}}$ such that $\mathbb{E}\{h(\Pi_1)\} \geq R$ and $\mathbb{E}\{f(\Pi_1)\} < g(R) = f \circ h^{-1}(R)$. Let $T = h(\Pi_1)$ and $\Pi_1 = h^{-1}(T)$. Then, $\mathbb{E}\{f \circ h^{-1}(T)\} = \mathbb{E}\{g(T)\} < g(R)$. Since g is a strictly convex function, $g(\mathbb{E}\{T\}) < \mathbb{E}\{g(T)\}$. In addition, as g is a monotone increasing function, $\mathbb{E}\{T\} < R$ must hold. However, $\mathbb{E}\{T\} = \mathbb{E}\{h(\Pi_1)\} < R$ results in a contradiction, which implies that the argument in the property holds, i.e., $p_{\Pi_1}^{\mathrm{opt}}(\pi_1) = \delta(\pi_1 - h^{-1}(R))$. ∎

Remark 5: Note that if $\gamma < \gamma_{th} \approx 0.166$, $g(r)$ is convex except over a short interval of low bit rates. Hence, in most of the practical scenarios, the result of Property 2 is expected to still hold.

IV. Numerical Results

In this section, numerical results are provided for the proposed signal constellation and/or prior distribution design problems. First, the optimal stochastic signaling is investigated

under average power and bit rate constraints based on the generic formulation in (8) and performance comparisons are conducted with respect to the alternative strategies proposed in Section III. In the examples, binary ($M = 2$) and quaternary ($M = 4$) communication systems with one dimensional observations ($n = 1$) are considered, and the following Gaussian mixture noise is employed:

$$
p_N(y) = \frac{1}{\sqrt{2\pi}\,\sigma L} \sum_{l=1}^{L} e^{-\frac{(y-\mu_l)^2}{2\sigma^2}}
\tag{28}
$$

where $L = 4$, $\mu_1 = -1.5$, $\mu_2 = -0.5$, $\mu_3 = 0.5$, and $\mu_4 = 1.5$.

The strategies evaluated in the examples are given below:

Optimal prior (deterministic): This strategy corresponds to the solution of (10). In this case, it is assumed that the constellation is fixed and the signals are specified as $\mathbf{s}_0 = -\sqrt{A}$ and $\mathbf{s}_1 = \sqrt{A}$ when $M = 2$. Note that for $M = 2$, the optimal prior distribution should satisfy the average bit rate constraint with equality according to Proposition 2. For $M = 4$, the fixed constellation signal points are specified as $\mathbf{s} = \left[\frac{-3}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{3}{\sqrt{5}}\right]$ with $A = 1$.

Optimal joint (deterministic): This strategy is obtained as the solution of (23), which yields the optimal deterministic prior probability and signal constellation vectors jointly.

Optimal joint (stochastic): This strategy corresponds to the solution of (8), which provides the optimal distribution for the prior probability and signal constellation vectors jointly.

In the first example, the binary signaling is used with $A = 1.2$ and $R = 0.8812 = h(0.3)$, and the average probability of error is calculated for various values of A/σ^2. It is observed from Fig. 1 that the jointly optimal stochastic design achieves the best performance, as expected, since it covers the other strategies as special cases. On the other hand, the optimal deterministic priors strategy yields the worst performance as it does not optimize the signal constellation vector together with the priors. The performance difference between various strategies becomes less significant in the low SNR regime. However, when $A/\sigma^2 > 12\,\mathrm{dB}$, one can notice the improvements over deterministic signaling via stochastic signaling.

Next, performance of the proposed strategies is investigated for $M = 4$. The power constraint is set as $A = 1$, and the same Gaussian mixture noise is employed as in the previous example. The average probabilities of error are calculated for the proposed strategies when $R = 1.9$ and $R = 2$. Recall that $R = 2$ corresponds to the use of equal priors for the constellation points. From Fig. 2, it is seen that employing a lower bit rate constraint improves the average probability of error performance for all the strategies. The best performance is again achieved via stochastic signaling, and the performance gap between the optimal joint stochastic signaling and the optimal joint deterministic signaling becomes larger for $R = 1.9$.

In order to observe behaviors of different strategies for varying bit rate constraints, SNR is fixed as $A/\sigma^2 = 24\,\mathrm{dB}$ and the average probabilities of error are plotted versus R. From Fig. 3, it is noted that the optimal joint stochastic and
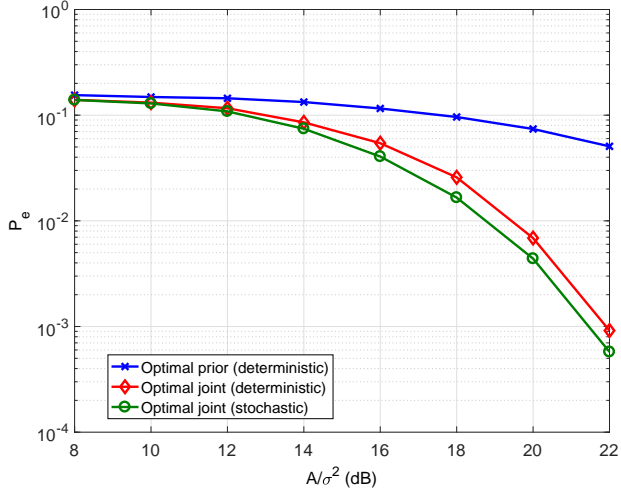
Fig. 1: P_e versus A/σ^2 for $M = 2$ with $A = 1.2$ and $R = 0.8812$ for different strategies.
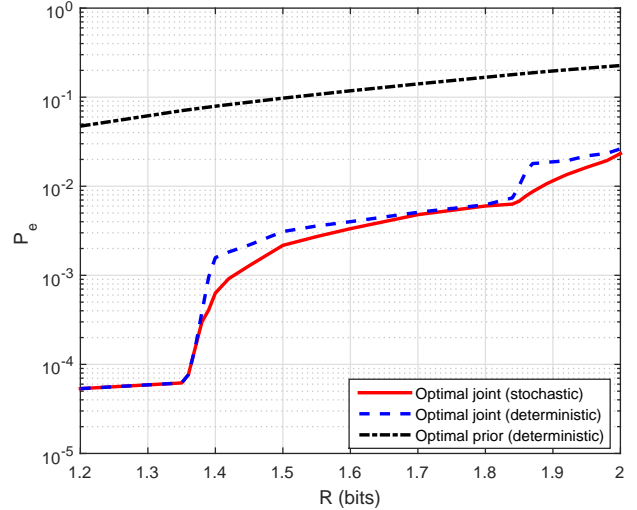

Fig. 3: P_e versus R for $M = 4$ under Gaussian mixture noise with $A = 1$ and $A/\sigma^2 = 24$ dB.
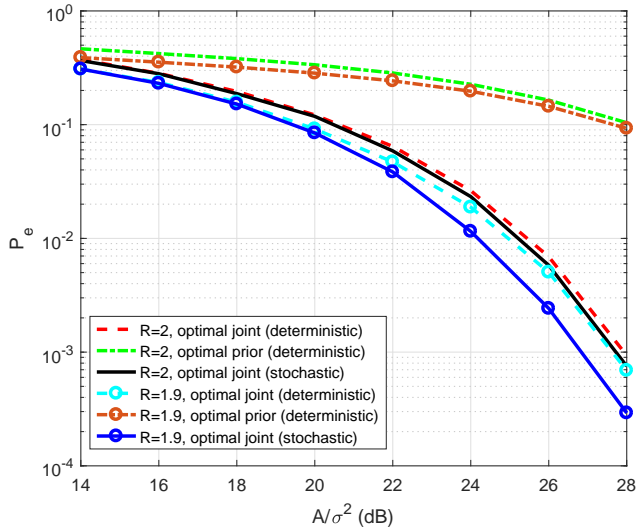

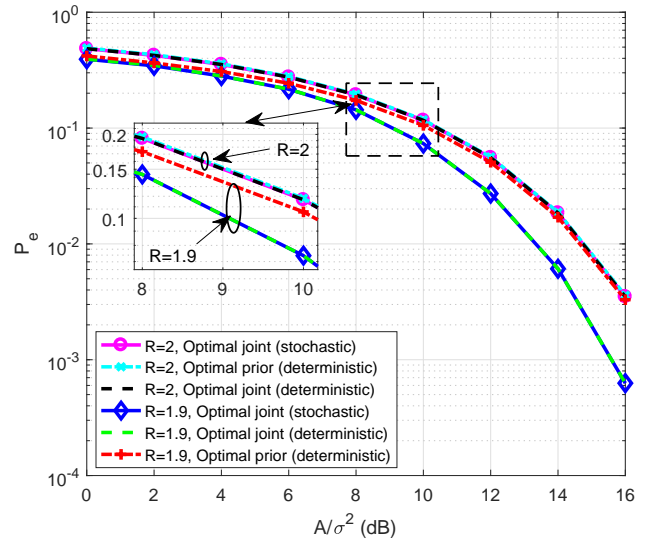Fig. 2: P_e versus A/σ^2 for $M = 4$ under Gaussian mixture noise with $A = 1$.

Fig. 4: P_e versus A/σ^2 for $M = 4$ under Gaussian noise with $A = 1$, $R = 2$ and $R = 1.9$.

deterministic approaches have the same solutions for low bit rate constraints ($R < 1.35$) and stochastic signaling improves the performance of deterministic signaling for medium and high R values as it allows randomization among different transmission policies (prior and signal constellation sets). Also, the sharp increase in the average probability of error around $R = 1.35$ and $R = 1.85$ is due to the fact that the effective noise has a multi-modal PDF.

Next, performance of the proposed strategies is investigated in the presence of zero-mean Gaussian noise for $M = 4$. From Fig. 4, it is observed for $R = 2$ that the optimal joint deterministic and stochastic solutions have the same performance (with the fixed constellation of $\mathbf{s} = \left[\frac{-3}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{3}{\sqrt{5}}\right]$), and the performance of the optimal prior solution is slightly worse. For $R = 1.9$, the optimal joint deterministic and stochastic approaches still achieve equal error probabilities, which are

significantly lower than those in the case of $R = 2$. On the other hand, the reductions in the error probabilities when R is reduced from 2 to 1.9 are very small for the optimal prior solution. This small performance difference reduces further as A/σ^2 increases.

Finally, we consider the 8-PAM modulation scheme to further evaluate the performance of the optimal deterministic prior design framework. The constellation is normalized to have unit average symbol power with respect to uniform signaling, i.e., $\Omega = \{\pm 1/\sqrt{21}, \pm 3/\sqrt{21}, \pm 5/\sqrt{21}, \pm 7/\sqrt{21}\}$. It is assumed that the received symbols are subject to zero-mean additive white Gaussian noise with variance σ^2, and consequently, the SNR is defined as $\text{SNR} = -10\log_{10}(\sigma^2)$. In Fig. 5, we depict the correct decision performance of the proposed optimal signaling scheme as a function of the
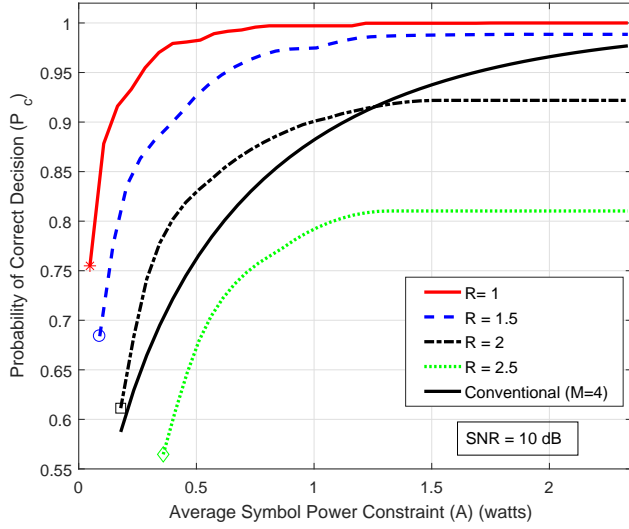
Fig. 5: Correct decision performance of the proposed method subject to constraints on the average bit rate and the average symbol power for an 8-point constellation $\Omega = \{\pm 1/\sqrt{21}, \pm 3/\sqrt{21}, \pm 5/\sqrt{21}, \pm 7/\sqrt{21}\}$.

constraint A on the average symbol power for different values of the average bit rate constraint $R \in \{1, 1.5, 2, 2.5\}$ when SNR = 10 dB. The marker shown at the leftmost end of each curve corresponds to the signal distribution that yields the minimum average symbol power under the specified constraint on the average bit rate. For each $R \in \{1, 1.5, 2, 2.5\}$, it is seen from the figure that the correct decision probability increases towards a limiting value as the constraint on the average symbol power is relaxed. Since we employ a fixed constellation, the maximum value of the correct decision probability is limited by the chosen value of average bit rate constraint even if the constraint on the average symbol power is large. Nevertheless, the proposed solution yields the optimal signaling distribution that maximizes the correct decision probability under constraints on the average bit rate and the average symbol power for the given constellation.

In order to compare the performance of the proposed scheme with that of the uniform signaling scheme, the correct decision probability of the conventional quaternary ($M = 4$) signaling with equally likely symbols is also depicted in Fig. 5 (see the solid black line). The conventional constellation for $M = 4$ is constructed as $\Omega_4(A) = \sqrt{A} \times \{\pm 1/\sqrt{5}, \pm 3/\sqrt{5}\}$ to yield an average symbol power value equal to A. It should be noted that as A increases, the minimum distance between the constellation points in $\Omega_4(A)$ increases, and hence, the correct decision probability improves steadily towards one. It is seen from Fig. 5 that the proposed approach (see the dash-dot black line corresponding to $R = 2$) yields higher correct decision performance with respect to the uniform quaternary signaling over the range $A \in (0.18, 1.28)$ while delivering an average bit rate of 2 bits per transmitted symbol. As an example, for $A = 0.51$, nonuniform signaling over the 8-point constellation Ω according to the signaling distribution

$\pi = (0.0588, 0.0084, 0.3043, 0.0186, 0.4272, 0.0143, 0.16424, 0.0041)$ attains a correct decision probability score approximately equal to 0.832 whereas the conventional uniform signaling over $\Omega_4(0.51)$ delivers 0.766. Hence, another advantage of the proposed scheme is that the correct decision performance can be improved with nonuniform signaling over a higher order constellation while satisfying the same average bit rate and average symbol power with those of a uniform signaling scheme over a lower order constellation.

V. CONCLUSION

In this paper, we have jointly optimized for the distribution of the signal constellation and the corresponding prior probability vector in order to minimize the average probability of error subject to constraints on average bit rate and average symbol power. Considering the prior probability vector as a part of the design leads to an extra degree of freedom compared to conventional stochastic signaling. Since the possible use of nonequal priors can reduce the average bit rate, we have imposed constraints on the average bit rate and power in the proposed formulation. The original formulation requires optimization over a space of joint PDFs, which is hard to solve in general. For this reason, we have first derived an alternative optimization problem, and proved that its solution achieves the same optimal value as that of the original problem. The advantage of the alternative formulation is that the optimal solution can be represented as a randomization among at most three different mass points; hence, it can be solved efficiently. After the general formulation, we have investigated three special cases focusing on the optimization of deterministic prior probabilities for a given fixed constellation, the optimal deterministic joint design of prior probabilities and constellation points and, a classical binary communication system with scalar observations under AWGN. Finally, numerical results have been presented for both the general formulation and the special cases.

A theoretical framework is presented in this paper for enhanced digital modulation by optimizing the prior probabilities and the corresponding signal constellation under average power and bit rate constraints. The idea of utilizing a flexible average bit rate (nonequal priors) to improve error performance can be applied to most digital communication systems, as the considered system model assumes a generic M-ary communication system under an additive noise channel, e.g., AWGN and flat fading channels with perfect channel estimation. Furthermore, the stochastic signaling approach can provide further improvements over deterministic signaling especially under additive non-Gaussian noise such as the Gaussian mixture. The effects of multiuser or co-channel interference and impulsive noise in communication systems can be modeled as Gaussian mixture noise [17]. Therefore, randomization of digital modulation can be an option to improve the average probability of error under such conditions. As a future work, we aim to extend this study to multi-user scenarios with varying average bit rate constraints and reliability targets, and design the modulation strategies in a non-orthogonal multiple access setting.
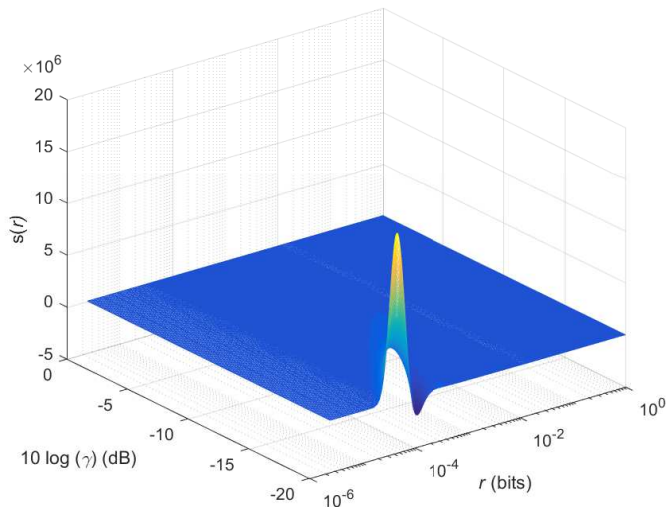
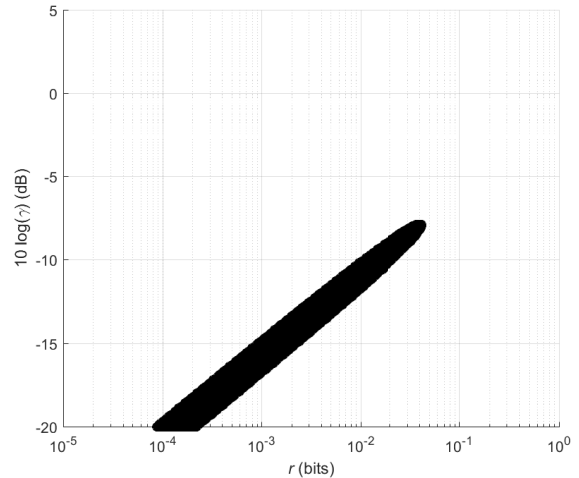Fig. 6: $s(r)$ versus r for various values of γ.

Fig. 7: The dark area shows the region in which $s(r) < 0$. Outside this region, $g(r)$ is convex.

APPENDIX B
DERIVATION OF PROPERTY 1

Let $\hat{h}(r) = h^{-1}(r) \in [0, 0.5]$. Then, $\frac{dg(r)}{dr} = f'(\hat{h}(r))\hat{h}'(r)$. Note that $\hat{h}(r)$ is a monotone increasing and convex function of r. It is first noted that $\frac{d^2 g(r)}{dr^2} = f''(\hat{h}(r))(\hat{h}'(r))^2 + f'(\hat{h}(r))(\hat{h}''(r))$. Since $h(\pi_1)$ is a one-to-one function on $\pi_1 \in [0, 0.5]$, we have $\hat{h}'(r) = 1/h'(\hat{h}(r))$. Hence, we obtain the following relation:

$$\frac{d^2 g(r)}{dr^2} = \frac{f''(\hat{h}(r))\, h'(\hat{h}(r)) - f'(\hat{h}(r))\, h''(\hat{h}(r))}{(h'(\hat{h}(r)))^3} \quad (31)$$

$$\triangleq \frac{s(r)}{(h'(\hat{h}(r)))^3}. \quad (32)$$

Note that the denominator of (31) is always positive as the binary entropy function is monotone increasing and concave on $[0, 0.5]$. Let $s(r)$ denote the numerator of (31). Then, the aim is to determine when $s(r) > 0$ to explore the convexity of $g(r)$. Fig. 6 shows $s(r)$ versus r for various γ settings. The numerical investigation reveals that $s(r)$ is positive for large values of γ; however, when $\gamma < \gamma_{th}$, it is negative in a certain interval of r values. This can be seen more clearly in Fig. 7, where the dark (black) region indicates the area in which $s(r) \le 0$. In addition, we provide Fig. 8 which illustrates $\frac{d^2 g(r)}{dr^2}$ for various values of γ. It is interesting to note that when $\gamma < \gamma_{th} \approx 0.166$, $g(r)$ is not convex for a certain interval of bit rates with small values.

APPENDIX C
PROOF OF LEMMA 2

In order to prove monotonicity, we need to show that $f'(\pi_1)$ is positive for $\pi_1 \in [0, 0.5]$. In the proof, $p \triangleq \pi_1$ and $\alpha \triangleq \sqrt{(1-p)/p}$ are used for convenience. By defining $u(p) \triangleq$
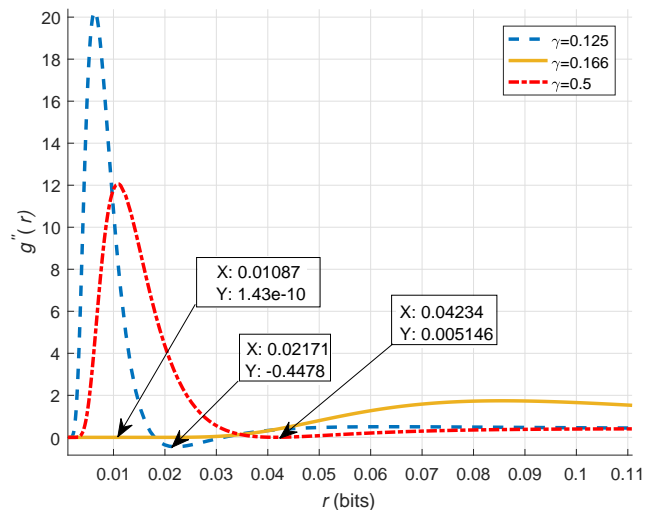

Fig. 8: $g''(r) = \frac{d^2 g(r)}{dr^2}$ versus r.

$\frac{\gamma}{2}(\alpha + \frac{1}{\alpha}) - \frac{\ln \alpha}{\frac{\gamma}{2}(\alpha + \frac{1}{\alpha})}$ and $v(p) \triangleq \frac{\gamma}{2}(\alpha + \frac{1}{\alpha}) + \frac{\ln \alpha}{\frac{\gamma}{2}(\alpha + \frac{1}{\alpha})}$, (27) can be rewritten as $f(p) = p\, Q(u(p)) + (1-p)\, Q(v(p))$. Then,

$$f'(p) = Q(u(p)) + p\, Q'(u(p))\, u'(p) - Q(v(p)) \\ + (1-p)\, Q'(v(p))\, v'(p). \quad (33)$$

In (33), $Q'(p) = -\frac{1}{\sqrt{2\pi}} e^{-p^2/2}$ and explicit formulas are required for $u'(\cdot)$ and $v'(\cdot)$. Note that $u'(p) = \frac{du}{d\alpha}\frac{d\alpha}{dp} = \tilde{u}(\alpha)\alpha'(p)$ by the chain rule. Similarly, $v'(p) = \frac{dv}{d\alpha}\frac{d\alpha}{dp} = \tilde{v}(\alpha)\alpha'(p)$, where $\alpha'(p) = \frac{-1}{2p\sqrt{p(1-p)}}$. Hence,

$$\tilde{u}(\alpha) = \frac{\gamma}{2}\left(1 - \frac{1}{\alpha^2}\right) - \frac{(1 + \frac{1}{\alpha^2}) - \ln \alpha\,(1 - \frac{1}{\alpha^2})}{\frac{\gamma}{2}(\alpha + \frac{1}{\alpha})^2} \quad (34)$$

and

$$\tilde{v}(\alpha) = \frac{\gamma}{2}\left(1 - \frac{1}{\alpha^2}\right) + \frac{(1 + \frac{1}{\alpha^2}) - \ln\alpha\,(1 - \frac{1}{\alpha^2})}{\frac{\gamma}{2}(\alpha + \frac{1}{\alpha})^2}. \quad (35)$$

For $p \in [0, 0.5)$,[2] we have $\alpha > 1$. Then, $u(p) < v(p)$ and $Q(u(p)) - Q(v(p)) > 0$ for any given γ, and $\alpha > 1$ as the Q-function is monotone decreasing. Thus, $f'(p)$ in (33) can be lower bounded as follows:

$$f'(p) > p\,Q'(u(p))\,u'(p) + (1 - p)\,Q'(v(p))\,v'(p) \quad (36)$$
$$= \frac{1}{\sqrt{8\pi\,p(1-p)}}\left(e^{-u(p)^2/2}\tilde{u}(\alpha) + e^{-v(p)^2/2}\tilde{v}(\alpha)\alpha^2\right).$$

Thus, it suffices to show that $e^{-u(p)^2/2}\tilde{u}(\alpha) + e^{-v(p)^2/2}\tilde{v}(\alpha)\alpha^2 > 0$. Then,

$$\frac{\gamma}{2}\left(\alpha + \frac{1}{\alpha}\right)^2\left(e^{-u(p)^2/2}\tilde{u}(\alpha) + e^{-v(p)^2/2}\tilde{v}(\alpha)\alpha^2\right)$$
$$= e^{\frac{u(p)^2}{-2}}\left(\frac{\gamma^2}{4}(1 - \frac{1}{\alpha^2})(\alpha + \frac{1}{\alpha})^2 + \ln\alpha(1 - \frac{1}{\alpha^2}) - (1 + \frac{1}{\alpha^2})\right)$$
$$+ e^{\frac{v(p)^2}{-2}}\left(\frac{\gamma^2}{4}(1 - \frac{1}{\alpha^2})(\alpha + \frac{1}{\alpha})^2\right.$$
$$\left. - \ln\alpha(1 - \frac{1}{\alpha^2}) + (1 + \frac{1}{\alpha^2})\right)\alpha^2 \triangleq c(\alpha). \quad (37)$$

Therefore, it is sufficient to determine if $c(\alpha)$ is positive. The $c(\alpha)$ term can be decomposed as

$$c(\alpha) = \left(e^{-\frac{u(p)^2}{2}} + \alpha^2 e^{-\frac{v(p)^2}{2}}\right)\left(1 - \frac{1}{\alpha^2}\right)\left(\frac{\gamma^2}{4}(\alpha + \frac{1}{\alpha})^2\right)$$
$$+ \left(e^{-u(p)^2/2} - \alpha^2 e^{-v(p)^2/2}\right)\left(\ln\alpha(1 - \frac{1}{\alpha^2}) - (1 + \frac{1}{\alpha^2})\right). \quad (38)$$

However, it can be shown that $e^{-u(p)^2/2} - \alpha^2 e^{-v(p)^2/2} = 0$ after inserting $u(p)$ and $v(p)$ into the expression. Furthermore, all the terms in the first line of (38) are strictly positive for any $\gamma > 0$ and $\alpha > 1$. This shows that $f'(p) > 0$ for $p \in [0, 0.5)$; hence it is a monotone increasing function.

REFERENCES

[1] B. Dulek, "Optimal signaling for a given constellation under average symbol energy and bit rate constraints," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Bilbao, Spain, Dec. 18-20 2017, pp. 112–116.

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

[3] I. Korn, J. Fonseca, and S. Xing, "Optimal binary communication with nonequal probabilities," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1435–1438, Sep. 2003.

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

[5] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1948–1960, Apr. 2010.

[6] A. B. Sediq, P. Djukic, H. Yanikomeroglu, and J. Zhang, "Optimized nonuniform constellation rearrangement for cooperative relaying," *IEEE Trans. Veh. Tech.*, vol. 60, no. 5, pp. 2340–2357, June 2011.

[7] N. S. Loghin, J. Zllner, B. Mouhouche, D. Ansorregui, J. Kim, and S. I. Park, "Non-uniform constellations for ATSC 3.0," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 197–203, March 2016.

[8] G. Foschini, R. Gitlin, and S. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Comm.*, vol. 22, no. 1, pp. 28–38, Jan. 1974.

[9] G. D. Forney, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 281–300, March 1992.

[10] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

[11] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. A. in Comm.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[12] G. D. Forney and L. F. Wei, "Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations," *IEEE J. Sel. A. Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[13] 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. Signal Process.*, vol. 55, no. 7, pp. 3172–3184, July 2007.

[14] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1655–1669, May 2009.

[15] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.

[16] B. Dulek and S. Gezici, "Detector randomization and stochastic signaling for minimum probability of error receivers," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 923–928, Apr. 2012.

[17] 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.

[18] ——, "Optimal signaling and detector design for power-constrained binary communications systems over non-Gaussian channels," *IEEE Commun. Lett.*, vol. 14, no. 2, pp. 100–102, Feb. 2010.

[19] ——, "On the optimality of stochastic signaling under an average power constraint," in *Proc. 48th Annual Allerton Conference on Communication, Control and Computing*, Monticello, Illinois, Sep. 29-Oct. 1 2010, pp. 1158–1164.

[20] M. E. Tutay, S. Gezici, and O. Arikan, "Optimal randomization of signal constellations on the downlink of multiuser DS-CDMA system," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 10, Oct. 2013.

[21] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[22] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[23] E. G. Larsson, "Improving the frame-error-rate of spatial multiplexing in block fading by randomly rotating the signal constellation," *IEEE Commun. Letters*, vol. 8, no. 8, pp. 514–516, Aug. 2004.

[24] ——, "Constellation randomization (CoRa) for outage performance improvement on MIMO channels," in *Proc. IEEE Global Telecommun. Conf.*, vol. 1, Nov.-Dec. 2004, pp. 386–390.

[25] C. Lamy and J. Boutros, "On random rotations diversity and minimum MSE decoding of lattices," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1584–1589, July 2000.

[26] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Tech.*, vol. 41, pp. 170–176, May 1992.

[27] X. Ma and G. G. B., "Space-time-multipath coding using digital phase sweeping," in *Proc. IEEE Global Comm. Conf.*, vol. 1, Nov. 2002, pp. 384–388.

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

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, USA: Wiley, 2006.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[31] D. R. Smart, *Fixed point theorems*, ser. Cambridge Tracts in Mathematics. Cambridge University Press Archive, 1980, vol. 66.

---

[2]Note that $f'(0.5) = 0$ but it does not effect the monotonicity if $f'(p) > 0$ for $p \in [0, 0.5)$.

Cagri Goken received the B.S. and M.S degrees from Bilkent University, Ankara, Turkey and the M.A. degree from Princeton University, NJ, USA, all in electrical engineering, in 2009, 2011 and 2014 respectively. He is currently working towards his Ph.D. degree at Bilkent University. He has also been with Aselsan Inc. since 2016, where he is currently a Senior Design Engineer. His research interests include detection and estimation theory, wireless communications and physical layer secrecy.

Berkan Dulek received the B.S., M.S. and Ph.D. degrees in electrical and electronics engineering from Bilkent University in 2003, 2006 and 2012, respectively. From 2007 to 2010, he worked at industry. From 2012 to 2013, he was a postdoctoral research associate at the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY. Since 2014, he has been with the Department of Electrical and Electronics Engineering at Hacettepe University, where he is currently an Associate Professor. His research interests are in statistical signal processing, detection and estimation theory, and communication theory. Dr. Dulek is an editor for Journal of Communications and Networks.

Sinan Gezici (S'03–M'06–SM'11) 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 2006 to 2007, he worked at Mitsubishi Electric Research Laboratories, Cambridge, MA. Since 2007, he has been with the Department of Electrical and Electronics Engineering at Bilkent University, where he is currently a Professor. Dr. Gezici's research interests are in the areas of detection and estimation theory, 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). Dr. Gezici was an associate editor for IEEE Transactions on Communications, IEEE Wireless Communications Letters, and Journal of Communications and Networks.