Optimal Joint Modulation Classification and Symbol Decoding

Ertan Kazikli, *Student Member, IEEE*, Berkan Dulek[ID], *Member, IEEE*, and Sinan Gezici[ID], *Senior Member, IEEE*

Abstract— In this paper, modulation classification and symbol decoding problems are jointly considered and optimal strategies are proposed under various settings. In the considered framework, there exist a number of candidate modulation formats and the aim is to decode a sequence of received signals with an unknown modulation scheme. To that aim, two different formulations are proposed. In the first formulation, the prior probabilities of the modulation schemes are assumed to be known and a formulation is proposed under the Bayesian framework. This formulation takes a constrained approach in which the objective function is related to symbol decoding performance whereas the constraint is related to modulation classification performance. The second formulation, on the other hand, addresses the case in which the prior probabilities of the modulation schemes are unknown, and provides a method under the minimax framework. In this case, a constrained approach is employed as well; however, the introduced performance metrics differ from those in the first formulation due to the absence of the prior probabilities of the modulation schemes. Finally, the performance of the proposed methods is illustrated through simulations. It is demonstrated that the proposed techniques improve the introduced symbol detection performance metrics via relaxing the constraint(s) on the modulation classification performance compared with the conventional techniques in a variety of system configurations.

Index Terms— Modulation classification, demodulation, Bayes, minimax.

I. INTRODUCTION

A. Motivation and Related Work

RECENT years have witnessed a surge of interest in modulation classification motivated by a variety of contemporary applications in which it is paramount to recognize the modulation format of a received signal prior to decoding the transmitted message [1]–[18]. In the classical signal reception scenario, it is assumed that the characteristics of the transmitted signal are known and the channel parameters can be estimated reliably at the receiver. Under this setting with perfect information regarding the modulation format employed by the transmitter, traditional digital receivers first perform several tasks such as downconversion, timing synchronization, matched filtering, sampling and channel equalization, and then demodulate the signal samples to recover the transmitted symbol sequence. If the modulation format employed by the transmitter is also unknown, the receiver needs to be equipped with a modulation classification mechanism before the symbol decoding stage [1]. For instance, in signal monitoring, a key topic in military electronic intelligence, there is no cooperation between the transmitting and the intercepting nodes to obtain information regarding the modulation format. Hence, in that case, it is imperative to perform modulation classification in order to convey the messages emitted by the target communications system. As another example, in cognitive and software defined radio based civilian applications, intelligent receivers equipped with modulation classification algorithms can be utilized to reduce the communication overhead resulting from the exchange of information regarding the modulation format adapted to the varying channel conditions [1], [4], [16].[1]

Modulation classification techniques can be grouped into feature-based and likelihood-based techniques [1], [3]. Feature-based techniques utilize features such as cumulants extracted from the received signal to identify the modulation format. Although feature-based techniques can be preferred due to their simplicity, these techniques are developed in an ad-hoc manner which do not guarantee optimality. On the other hand, likelihood-based techniques are developed using a statistical characterization of the received signal which facilitates the development of optimal (or near optimal in case the analysis uses some approximations, simplifying assumptions, etc.) techniques. In this study, in order to come up with optimal techniques for the solution of the considered problem, a statistical characterization of the received signal is utilized, and hence, related literature focusing on likelihood-based techniques are summarized in the following. In the context of likelihood-based modulation classification techniques, three prominent approaches are adopted based on how the unknown channel parameters and the transmitted symbol sequence are handled in the statistical model: the average likelihood ratio test (ALRT) [1], [2], [5], [8], [10], [15], [18], the generalized likelihood ratio test (GLRT) [1], [2], [5], [13], [17], [20] and the hybrid likelihood ratio test (HLRT) [1], [2], [5], [9], [10],

[1]In adaptive modulation, the transmitter employs a low order modulation scheme under low signal-to-noise ratio (SNR) in order to ensure reliable transmission of the data whereas a high order modulation scheme is employed to increase the data rate in the case of high SNR [19].

[13], [14], [18], [21]. In the case of ALRT, all unknowns are treated as random variables with known prior distributions so that the likelihood expressions can be marginalized over the space of channel parameters and modulation symbols. Then, the resulting average likelihood scores are compared to decide on the modulation format. On the other hand, GLRT treats all unknowns as deterministic and substitutes their maximum likelihood (ML) estimates in the corresponding likelihood expressions to obtain a likelihood ratio based decision metric. HLRT takes a compromise between ALRT and GLRT by treating some parameters as random quantities with known priors while considering the rest as deterministic.

It should be noted that the primary goal of the aforementioned techniques is to improve the modulation classification accuracy. It is sensible to assume that better modulation classification performance yields better symbol decoding performance, the latter being the ultimate task for any receiver equipped with a modulation classifier. However, since these approaches have not been optimized for the subsequent symbol decoding stage, it is also reasonable to question their optimality. In this work, it is shown that by jointly designing the modulation classification and the symbol decoding stages, significant performance improvements for the symbol detection performance over the conventional strategy that treats two problems separately can be achieved.

Our perspective on the problem has been motivated by the success of joint design frameworks such as joint detection and estimation [22]–[27] and joint detection and decoding [28]–[30]. The work in [23] considers a binary hypothesis testing problem in which only one of the hypothesis is composite, i.e., it contains an unknown random parameter whose prior distribution is assumed to be known. Under this setting, the aim is to jointly determine the correct hypothesis and to provide an estimate of the unknown parameter when a decision is declared in favor of the composite hypothesis. It is shown that the optimal decision rule depends on both the likelihood ratio and the posterior estimation cost for the unknown parameter while the optimal parameter estimate is given by the classical Bayes estimator that minimizes the posterior estimation cost. In [24], the optimal estimators and decision rule are determined in order to minimize the maximum of the conditional estimation costs under constraints on the false alarm and miss detection probabilities. In this case, unknown random parameters are assumed under both hypotheses. In [25], the authors consider the average (Bayesian) estimation cost as the objective function. Different from the previous work, unknown parameters common to both hypotheses are assumed in addition to parameters unique to each hypothesis. The optimization problems considered in [23]–[25] are formulated in a similar manner in the sense that the estimation performance metric appears in the objective as a function of both the estimators and the decision rule whereas the constraints are related to the detection performance metric, being solely dependent on the decision rule (i.e., uncoupled with the estimator), which facilitates ease of analysis. In [28], the authors consider intermittent communication over a noisy channel. It is assumed that the receiver acquires either pure noise or a codeword and the aim is to determine the

presence of a codeword and to decode if declared present. The optimal decision rule that minimizes the probability of not correctly deciding on the true codeword transmitted subject to constraints on the false alarm and miss detection probabilities of rejecting pure noise hypothesis is derived. In [29], this idea is extended to channel detection in coded communications.

In addition, modulation classification and symbol detection problems are jointly considered for multiple-input multiple-output (MIMO) systems in [12] and [31]–[33], which aim to design low complexity algorithms. On the other hand, in this paper, we focus on the development of *optimal* algorithms for this joint design problem considering traditional single-input-single-output (SISO) systems.

B. Main Contributions

The main contributions of this paper can be summarized as follows:

- For the first time in the literature, modulation classification and symbol decoding problems are considered in a joint framework. Both the Bayesian and the minimax formulations are provided depending on the availability of the prior probabilities for the employed modulation schemes. In the presence of prior information, the optimal decision rule that minimizes the overall average probability of symbol error (at the subsequent stage following modulation classification) subject to a constraint on the average probability of erroneous modulation classification is obtained in closed form. In the absence of such information, the optimal decision rule that minimizes the maximum symbol error probability given that the modulation is correctly identified is derived under individual constraints on the conditional correct classification probability of each modulation scheme.
- Since the proposed modulation classification rules are optimal with respect to each criterion, no other classifier can achieve a lower average (or maximum) symbol error probability while delivering equal or smaller average (or individual) modulation classification error(s).
- The proposed framework provides an optimal tradeoff between the probabilities of symbol error and modulation classification error. It is general in the sense that it covers the conventional method of performing LRT based modulation classification followed by ML decoding of the transmitted symbols as a special case.

C. Organization of the Paper

This paper is organized as follows. Section II presents the system model for the problem of modulation classification and symbol decoding as well as the conventional technique that relies on separate treatment. Section III-A formulates the problem of joint modulation classification and symbol decoding for the case of known prior probabilities of the modulation scheme and derives the optimal modulation classification and symbol decoding rules. Next, Section III-B considers a minimax formulation of the joint problem in the absence of modulation scheme priors. Section III-C presents some remarks on the proposed approach. Section IV discusses the computational

complexity of the proposed techniques as well as that of the conventional techniques. Section V describes an extension of the proposed framework obtained by slightly altering the metric used to measure the symbol decoding performance. Section VI provides numerical examples to illustrate the performance of the proposed solutions and their comparison with the conventional approach. Section VII concludes the paper with some remarks.

II. PROBLEM FORMULATION

A. System Model

Consider a coherent and synchronous signal reception scenario under which the discrete-time baseband signal model is given by [34]–[37]

$$x_k = s_k + \eta_k, \quad k = 1, \ldots, K, \tag{1}$$

where s_k and η_k are the transmitted constellation symbol and the additive noise term at time k, respectively, and K denotes the number of observations. Suppose there are L candidate digital amplitude-phase modulation schemes and let the set of constellation points corresponding to the ℓth modulation scheme be denoted as $\mathcal{S}^{(\ell)} \triangleq \{\mu_1^{(\ell)}, \ldots, \mu_{M^{(\ell)}}^{(\ell)}\}$ for $\ell = 1, \ldots, L$, where $\mu_m^{(\ell)}$ represents a certain amplitude and/or phase modulated complex signal corresponding to the mth symbol of the ℓth modulation scheme and $M^{(\ell)}$ denotes the modulation order of the ℓth modulation scheme. It is assumed that during the reception of K signal samples, the modulation scheme remains the same, i.e., $(s_k)_{k=1}^K \in \mathcal{S}^{(\ell)}$ for some fixed ℓ. Suppose that under each modulation scheme the symbols are equally likely and the symbols transmitted at distinct time instants are independent. Furthermore, let $(\eta_k)_{k=1}^K$ be independent and identically distributed (i.i.d.) according to a known probability density function (PDF) denoted by $f_\eta^{(\ell)}(\cdot)$ when the ℓth modulation scheme is employed. By allowing different PDFs for the noise term under each modulation scheme, the effect of different channel conditions such as that of SNR variations in an adaptive modulation system are incorporated into the model given in (1). The conditional PDF of the observed signal sample x_k given that the mth symbol of the ℓth modulation scheme is sent at time instant k is expressed as $f^{(\ell)}(x_k \mid \mu_m^{(\ell)}) \triangleq f_\eta^{(\ell)}(x_k - \mu_m^{(\ell)}).$[2]

Given the observed samples $\boldsymbol{x} \triangleq [x_1, \ldots, x_K]^T$, the aim is to identify the modulation scheme $\mathcal{S}$ and to decode the transmitted symbol sequence $\boldsymbol{s} \triangleq [s_1, \ldots, s_K]^T$. We consider two types of decision functions: $\boldsymbol{\delta} \triangleq \{\delta^{(\ell)}(\cdot)\}_{\ell=1}^L$ and $\boldsymbol{\phi} \triangleq \{\phi^{(\ell)}(\cdot)\}_{\ell=1}^L$. $\delta^{(\ell)}(\boldsymbol{x})$ denotes the probability of choosing the ℓth modulation scheme based on the observation $\boldsymbol{x}$ such that $\sum_{\ell=1}^L \delta^{(\ell)}(\boldsymbol{x}) = 1$ and $\delta^{(\ell)}(\boldsymbol{x}) \geq 0$ for every ℓ and $\boldsymbol{x}$. On the other hand, $\phi^{(\ell)}(x)$ represents the decision rule for decoding transmitted symbols under the ℓth modulation scheme based on the observed sample x. More specifically, $\phi^{(\ell)}(x_k) = m$ for some $m \in \{1, \ldots, M^{(\ell)}\}$ means that the transmitted symbol s_k is decoded as $\mu_m^{(\ell)}$. Whenever a decision is declared in

favor of the ℓth modulation scheme, all samples are decoded using $\phi^{(\ell)}(\cdot)$.

Before proceeding with the proposed framework for joint modulation classification and symbol decoding, we summarize the conventional approach that treats two problems separately along with their solutions.

B. Conventional Method

1) Modulation Classification: Identifying the modulation scheme based on the samples given in (1) can be accomplished in two different ways depending on the availability of the prior probabilities for the modulation schemes under consideration. For example, in an adaptive modulation system that employs different constellations for low and high SNR regimes, the proportion of time that the channel spends in each SNR regime can be employed in the assessment of prior probabilities. If modulation type priors are known, the Bayes classifier is optimal [20]. Since the received samples are i.i.d., the likelihood function (LF) of the ℓth modulation scheme is written as

$$f^{(\ell)}(\boldsymbol{x}) = \prod_{k=1}^K f^{(\ell)}(x_k), \tag{2}$$

where

$$f^{(\ell)}(x_k) \triangleq \sum_{m=1}^{M^{(\ell)}} \frac{1}{M^{(\ell)}} f^{(\ell)}(x_k \mid \mu_m^{(\ell)}). \tag{3}$$

Let $\boldsymbol{c} \triangleq [c_1, \ldots, c_L]$, where c_ℓ denotes the prior probability of the ℓth modulation scheme. Then, the probability of classification error is minimized by the maximum *a posteriori* (MAP) decision rule [20]:

$$\delta_{\mathrm{B}}^{(j)}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } j = \arg\max_{\ell \in \{1, \ldots, L\}} c_\ell f^{(\ell)}(\boldsymbol{x}) \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The Bayes classifier (i.e., the MAP classifier) in (4) reduces to the ML classifier when equal priors for modulation schemes are assumed. On the other hand, if the prior probabilities of the modulation schemes are unknown, Neyman Pearson (NP) framework can be employed for classification between two candidate modulation schemes.[3] In this case, two types of classification errors are given by $P_1(\delta^{(1)}, \delta^{(2)}) \triangleq \int \delta^{(2)}(\boldsymbol{x}) f^{(1)}(\boldsymbol{x}) d\boldsymbol{x}$ and $P_2(\delta^{(1)}, \delta^{(2)}) \triangleq \int \delta^{(1)}(\boldsymbol{x}) f^{(2)}(\boldsymbol{x}) d\boldsymbol{x}$, where $\delta^{(1)}(\boldsymbol{x}) \geq 0$, $\delta^{(2)}(\boldsymbol{x}) \geq 0$ and $\delta^{(1)}(\boldsymbol{x}) + \delta^{(2)}(\boldsymbol{x}) = 1$ are satisfied for all $\boldsymbol{x}$. In the NP framework, the decision rule that minimizes one type of error subject to a constraint on the other is desired. Due to asymmetry between the objective and the constraint, we denote by $\{\delta_{\mathrm{NP},i}^{(\ell)}(\cdot)\}_{\ell=1}^2$ the NP type modulation classifier that places the constraint on the ith type of error, i.e., $P_i(\delta^{(1)}, \delta^{(2)}) \leq \beta_i$, and minimizes the other type of error. Based on the NP lemma [20], the corresponding modulation classifier is given by the following LRT:

$$\delta_{\mathrm{NP},i}^{(2)}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \dfrac{f^{(2)}(\boldsymbol{x})}{f^{(1)}(\boldsymbol{x})} > \tau_i \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

---

[2]Although an additive noise channel is assumed in this paper, the proposed framework generalizes in a straightforward manner to other memoryless channel models thanks to the general form of the statistical characterization $f^{(\ell)}(x_k \mid \mu_m^{(\ell)})$.

[3]In the case of more than two modulation schemes, a generalization of NP lemma to more than one side condition can be utilized [38, Sec. 3.6].

with $\delta_{\mathrm{NP},i}^{(1)}(\boldsymbol{x}) = 1 - \delta_{\mathrm{NP},i}^{(2)}(\boldsymbol{x})$ for $i = 1, 2$ where τ_i is chosen to satisfy $P_i(\delta_{\mathrm{NP},i}^{(1)}, \delta_{\mathrm{NP},i}^{(2)}) = \beta_i$. It should be noted that if the likelihood ratio in (5) contains any point mass, a randomization may be required in order to satisfy the constraint with equality.

Another alternative for modulation classification in the absence of modulation scheme prior probabilities is to employ the minimax hypothesis testing framework in which the maximum of individual probabilities of classification error, i.e., $\max_{\ell \in \{1,\ldots,L\}} 1 - \int \boldsymbol{\delta}^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}$, is minimized [20]. Under this framework, the modulation classification rule is a Bayes rule that corresponds to a least favorable prior.

2) Symbol Decoding: Once a decision is declared in favor of a modulation scheme, the MAP decision rule, which selects the symbol with the maximum *a posteriori* probability, can be employed to minimize the probability of symbol error for the corresponding constellation [19]. The *a posteriori* probability that the mth symbol of the ℓth modulation scheme is transmitted given the observed sample x_k is computed as

$$p^{(\ell)}(\mu_m^{(\ell)} \mid x_k) \triangleq \frac{1}{M^{(\ell)}} \frac{f^{(\ell)}(x_k \mid \mu_m^{(\ell)})}{f^{(\ell)}(x_k)} \qquad (6)$$

and the MAP symbol decoding rule for the ℓth modulation scheme is expressed as

$$\phi_{\mathrm{MAP}}^{(\ell)}(x_k) = \arg\max_{m \in \{1,\ldots,M^{(\ell)}\}} p^{(\ell)}(\mu_m^{(\ell)} \mid x_k). \qquad (7)$$

To sum up, in the conventional approach, the modulation scheme is determined first based on either (4) or (5) (or, the Bayes rule for least favorable prior when the minimax framework is adopted) depending on the availability of the prior probabilities of the modulation schemes and then symbols are decoded according to (7) for the selected modulation.

III. OPTIMAL JOINT MODULATION CLASSIFICATION AND SYMBOL DECODING

A. Known Priors for the Modulation Schemes

In this part, a joint formulation of the modulation classification and the symbol decoding problems is proposed for the case when prior probabilities of the modulation schemes are known or accurately estimated. A constrained optimization framework is considered in which the objective function reflects our desire to accurately recover as much symbols as possible while the constraint function controls the probability of classification error for the modulation type. In particular, consider the following cost function related to symbol error rate:

$$\mathrm{C}(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m}) \triangleq \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{\{\phi^{(\ell)}(x_k) \neq m_k\}} \qquad (8)$$

where $\boldsymbol{m} \triangleq [m_1, \ldots, m_K]^T$ contains the indices of the actual transmitted symbols from the ℓth modulation scheme with $m_k \in \{1, \ldots, M^{(\ell)}\}$ and $\mathbb{1}_{\{\cdot\}}$ is the indicator function that returns 1 if the condition in its argument is satisfied and 0 otherwise. Our aim is to minimize the average cost of symbol error when the employed modulation scheme is correctly identified while controlling the probability of modulation classification error (and consequently, the erroneous symbol

decoding due to an incorrect modulation decision). To that aim, the average cost of symbol error in the presence of a correct modulation classification decision can be obtained by taking the expectation of (8) over the joint PDF of transmitted symbols and received samples when the employed modulation scheme is correctly identified and averaging over all candidate modulation schemes as follows:

$$
\begin{aligned}
\mathcal{J}_{\mathrm{a}}(\boldsymbol{\delta}, \boldsymbol{\phi}) &\triangleq \sum_{\ell=1}^{L} c_\ell \, \mathrm{E}_\ell[\delta^{(\ell)}(\boldsymbol{x}) \, \mathrm{C}(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})] \\
&= \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) \sum_{m_1=1}^{M^{(\ell)}} \cdots \sum_{m_K=1}^{M^{(\ell)}} \frac{1}{K} \\
&\quad \sum_{k=1}^{K} \mathbb{1}_{\{\phi^{(\ell)}(x_k) \neq m_k\}} \left(\prod_{k=1}^{K} \frac{f^{(\ell)}(x_k \mid \mu_{m_k}^{(\ell)})}{M^{(\ell)}} \right) d\boldsymbol{x},
\end{aligned}
$$
$$(9)$$

where $\mathrm{E}_\ell[\cdot]$ denotes expectation with respect to (w.r.t.) the conditional joint PDF of transmitted symbols and received samples given that the ℓth modulation scheme is employed and the subscript a in $\mathcal{J}_{\mathrm{a}}(\cdot)$ indicates averaging over the candidate modulation schemes. The aim is to minimize the objective function proposed in (9) with respect to $\{\delta^{(\ell)}(\cdot)\}_{\ell=1}^{L}$ and $\{\phi^{(\ell)}(\cdot)\}_{\ell=1}^{L}$. Notice that the cost function given in (9) takes into account the symbol decoding error when the modulation is correctly recognized. On the other hand, if the modulation is not correctly identified, the symbol decoding stage uses the wrong set of constellation symbols while recovering the transmitted symbol sequence, i.e., although $(s_k)_{k=1}^{K} \in \mathcal{S}^{(\ell)}$ in (1), $\phi^{(\ell')}(\cdot)$ is utilized to decode symbols for some $\ell' \neq \ell$. As a result, symbol decoding performance in the case of incorrect modulation classification becomes significantly poor.[4] In that respect, there is also a need to control the probability of modulation classification error as it causes erroneous symbol decoding. For fixed modulation classification rule $\boldsymbol{\delta}$, the average probability of classification error is given by

$$\mathcal{H}_{\mathrm{a}}(\boldsymbol{\delta}) = 1 - \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}, \qquad (10)$$

which is seen to be independent of the symbol decoders $\boldsymbol{\phi}$. Subsequently, a constraint on the average probability of modulation classification error is introduced, which, together with (9), yields the following optimization problem:

$$
\begin{aligned}
&\min_{\boldsymbol{\delta}, \boldsymbol{\phi}} \quad \mathcal{J}_{\mathrm{a}}(\boldsymbol{\delta}, \boldsymbol{\phi}) \\
&\text{subject to} \quad \mathcal{H}_{\mathrm{a}}(\boldsymbol{\delta}) \leq \alpha,
\end{aligned}
\qquad (\mathrm{P1})
$$

where α is the constraint on the average probability of modulation classification error. Let α_{B} denote the average probability of modulation classification error corresponding to the Bayes rule given in (4). Since no detector can yield lower probability of modulation classification error than that

---

[4]A modulation classification error does not always mean that all the symbols are decoded erroneously (e.g., when some constellation points from different modulation schemes coincide). However, since the mapping of bits onto symbols will be different for distinct constellations, a modulation classification error, in general, significantly degrades the decoding performance.

of the Bayes detector, it is required that $\alpha \geq \alpha_B$ is satisfied for feasibility of the optimization problem. More generally, α can be regarded as a design parameter that governs the tradeoff between modulation classification performance and symbol decoding performance.

In order to obtain the solution of the optimization problem given in (P1), we first present the following lemma that expresses optimality of the MAP rule based symbol decoding.

Lemma 1: For any given priors c and feasible modulation classification rule specified by δ, the optimal symbol decoders that minimize $\mathcal{J}_a(\delta, \phi)$ are obtained as the MAP decision rules $\phi_{MAP} \triangleq \{\phi_{MAP}^{(\ell)}(\cdot)\}_{\ell=1}^L$ given in (7).

Proof: Since the modulation classification rule $\delta = \{\delta^\ell(\cdot)\}_{\ell=1}^L$ is fixed and the constraint function does not depend on $\phi = \{\phi^{(\ell)}(\cdot)\}_{\ell=1}^L$, (P1) reduces to an unconstrained problem in this case. Moreover, the problem decouples across ℓ since we minimize the summation of L terms each of which depend only on $\phi^{(\ell)}(\cdot)$. The optimization problem related to the symbol decoder under the ℓth modulation scheme is expressed as

$$\min_{\phi^{(\ell)}} \int \delta^{(\ell)}(\boldsymbol{x}) \sum_{m_1=1}^{M^{(\ell)}} \cdots \sum_{m_K=1}^{M^{(\ell)}} \left(\sum_{k=1}^K \mathbb{1}_{\{\phi^{(\ell)}(x_k) \neq m_k\}} \right)$$
$$\left(\prod_{k=1}^K \frac{1}{M^{(\ell)}} f^{(\ell)}(x_k \mid \mu_{m_k}^{(\ell)}) \right) d\boldsymbol{x}. \quad (11)$$

By noting that $\mathbb{1}_{\{\phi^{(\ell)}(x_k) \neq m_k\}} = 1 - \mathbb{1}_{\{\phi^{(\ell)}(x_k) = m_k\}}$ and using (2), (3), and (6), the optimization problem in (11) can be written as

$$\max_{\phi^{(\ell)}} \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \sum_{m_1=1}^{M^{(\ell)}} \cdots \sum_{m_K=1}^{M^{(\ell)}} \left(\sum_{k=1}^K \mathbb{1}_{\{\phi^{(\ell)}(x_k) = m_k\}} \right)$$
$$\left(\prod_{k=1}^K p^{(\ell)}(\mu_{m_k}^{(\ell)} \mid x_k) \right) d\boldsymbol{x}, \quad (12)$$

which after some manipulations reduces to

$$\max_{\phi^{(\ell)}} \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \sum_{k=1}^K \sum_{m_k=1}^{M^{(\ell)}} \mathbb{1}_{\{\phi^{(\ell)}(x_k) = m_k\}}$$
$$p^{(\ell)}(\mu_{m_k}^{(\ell)} \mid x_k) d\boldsymbol{x}. \quad (13)$$

From (13), it is seen that the objective function is maximized by selecting the symbol with the maximum *a posteriori* probability $p^{(\ell)}(\mu_{m_k}^{(\ell)} \mid x_k)$, which corresponds to the MAP decoder given in (7). ∎

Lemma 1 states that the MAP symbol decoder is optimal irrespective of the employed modulation classification rule. From a practical perspective, it means that there is no need to change the existing receiver demodulator architecture. On the other hand, as it will be shown next, the proposed joint formulation leads to a modulation classification rule that is different from the conventional Bayes classifier given in (4).

Substituting the optimal MAP symbol decoders into the objective function of (P1), we get

$$\mathcal{J}_a(\delta, \phi_{MAP}) = \sum_{\ell=1}^L c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) C_p^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}. \quad (14)$$

In (14), $C_p^{(\ell)}(\boldsymbol{x})$ is the posterior cost associated with the optimal MAP based symbol decoder for the ℓth constellation, which is given by

$$C_p^{(\ell)}(\boldsymbol{x}) = 1 - \frac{1}{K} \sum_{k=1}^K \max_{m \in \{1, \ldots, M^{(\ell)}\}} p^{(\ell)}(\mu_m^{(\ell)} \mid x_k). \quad (15)$$

Consequently, the optimization problem in (P1) is obtained as

$$\min_{\delta} \quad \sum_{\ell=1}^L c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) C_p^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}$$

$$\text{subject to} \quad 1 - \sum_{\ell=1}^L c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \alpha. \quad (16)$$

Next, the following theorem provides the optimal modulation classification rule that solves (16), which together with Lemma 1, characterize the solution to the original optimization problem (P1).

Theorem 1: For $\alpha \geq \alpha_B$, and assuming that $\{f^{(\ell)}(\boldsymbol{x}) (C_p^{(\ell)}(\boldsymbol{x}) - \lambda)\}_{\ell=1}^L$ do not contain any nonzero probability mass for $\lambda \geq 0$, the optimal modulation classification rule that solves (16) is given by

Case 1:

$$\delta^{(j)}(\boldsymbol{x}) = 1$$
$$\text{if } j = \arg\min_{\ell \in \{1, \ldots, L\}} c_\ell f^{(\ell)}(\boldsymbol{x}) C_p^{(\ell)}(\boldsymbol{x}) \quad (17)$$

and $\delta^{(j)}(\boldsymbol{x}) = 0$ otherwise, for $j = 1, \ldots, L$, if the constraint on the classification error probability is satisfied, i.e., $\mathcal{H}_a(\delta) \leq \alpha$; if not go to Case 2.

Case 2:

$$\delta^{(j)}(\boldsymbol{x}) = 1$$
$$\text{if } j = \arg\min_{\ell \in \{1, \ldots, L\}} c_\ell f^{(\ell)}(\boldsymbol{x}) \left(C_p^{(\ell)}(\boldsymbol{x}) - \lambda \right) \quad (18)$$

and $\delta^{(j)}(\boldsymbol{x}) = 0$ otherwise, for $j = 1, \ldots, L$, where $\lambda > 0$ is selected such that the probability of classification error is equal to α, i.e., $\mathcal{H}_a(\delta) = \alpha$.

Proof: The optimization problem given in (16) is linear and hence convex in $\{\delta^{(\ell)}(\cdot)\}_{\ell=1}^L$. Moreover, with the assumption of $\alpha \geq \alpha_B$, the problem is feasible and Slater's condition is satisfied. Therefore, Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality. We begin with writing the Lagrangian as follows:

$$\mathcal{L}(\delta, \lambda) = \lambda(1 - \alpha)$$
$$+ \int \sum_{\ell=1}^L \delta^{(\ell)}(\boldsymbol{x}) c_\ell f^{(\ell)}(\boldsymbol{x}) \left(C_p^{(\ell)}(\boldsymbol{x}) - \lambda \right) d\boldsymbol{x} \quad (19)$$

Using the fact that $\delta^{(\ell)}(\boldsymbol{x}) \geq 0$ for all ℓ and $\sum_\ell \delta^{(\ell)}(\boldsymbol{x}) = 1$, it is seen that (19) is minimized when modulation classifier selects the modulation that minimizes $c_\ell f^{(\ell)}(\boldsymbol{x})(C_p^{(\ell)}(\boldsymbol{x}) - \lambda)$ for each given $\boldsymbol{x}$. Based on the remaining KKT conditions, λ can be determined. Namely, dual feasibility, complementary slackness and primal feasibility conditions require that $\lambda \geq 0$, $\lambda(1 - \alpha - \sum_{\ell=1}^L c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}) = 0$ and $1 - \sum_{\ell=1}^L c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \alpha$, respectively. First, let

us substitute $\lambda = 0$. If the primal feasibility condition is satisfied, then the corresponding classifier given in (17) is the optimal solution to (16). On the other hand, if the classifier given in (17) violates the constraint, we go to Case 2, where $\lambda > 0$ is assumed. In this case, modulation classification constraint must be satisfied with equality due to complementary slackness. As $\lambda \to \infty$, it is seen from (18) that the optimal decision rule converges to the conventional MAP decision rule given in (4), which yields the lowest modulation classification error, i.e., $\mathcal{H}_a(\boldsymbol{\delta}_B) = \alpha_B$, among all detectors. Noting that the probability of modulation classification error corresponding to the detector given in (18) is a continuous function of λ, by the intermediate value theorem, there exists $\lambda > 0$ such that $\mathcal{H}_a(\boldsymbol{\delta}(\lambda)) = \alpha$. Hence, a solution in the form of (18) that satisfies all KKT conditions is guaranteed. ∎

It is seen from (17) and (18) that the optimal modulation classification rule takes into account *a posteriori* probabilities of the transmitted symbols in addition to the likelihood of the observed samples and the prior probabilities of the modulation schemes. As a result, symbol decoding performance is taken into account in the proposed modulation classification mechanism. For a given feasible constraint on the modulation classification error, no other mechanism can achieve better average symbol decoding performance when the modulation type is correctly identified. In that respect, the proposed framework provides the flexibility of tradeoff between performances of modulation classification and symbol decoding. Furthermore, the conventional method, which treats two problems separately, is obtained as a special case of the proposed method by selecting $\alpha = \alpha_B$. By relaxing the constraint on the modulation classification performance, i.e., taking $\alpha > \alpha_B$, it is possible to improve the average symbol decoding performance in the case of correct modulation classification. To sum up, the solution to the proposed joint optimization problem stated in (P1) is given by the modulation classifier specified in (17) and (18), and the standard MAP rule based symbol decoder corresponding to the classified modulation scheme which is given in (7).

B. Unknown Priors for the Modulation Schemes

The approach proposed in the previous part requires that the prior probabilities of the modulation schemes are known or accurately estimated. In order to come up with a robust joint scheme that can correctly identify the modulation scheme in the absence of this information and decode the received samples, we consider the maximum conditional symbol decoding error among candidate modulation schemes given that each is identified correctly as our performance metric. More explicitly, the objective function is expressed as

$$\mathcal{J}_m(\boldsymbol{\delta}, \boldsymbol{\phi}) \triangleq \max_{\ell \in \{1,\dots,L\}} E_\ell[C(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m}) \mid D = \mathcal{S}^{(\ell)}]$$
$$= \max_{\ell \in \{1,\dots,L\}} \frac{E_\ell[C(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})\delta^{(\ell)}(\boldsymbol{x})]}{E_\ell[\delta^{(\ell)}(\boldsymbol{x})]}, \quad (20)$$

where the subscript m in $\mathcal{J}_m$ denotes that maximum is taken over the modulation schemes, $\{D = \mathcal{S}^{(\ell)}\}$ denotes the event that a decision is declared in favor of the ℓth modulation

scheme, and the subscript ℓ denotes that the ℓth modulation scheme is employed. As for the constraints, since an average probability of classification error cannot be defined in this case, an individual constraint on each conditional probability of classification error is employed. This results in the following optimization problem:

$$\min_{\boldsymbol{\delta}, \boldsymbol{\phi}} \quad \mathcal{J}_m(\boldsymbol{\delta}, \boldsymbol{\phi})$$
$$\text{subject to} \quad 1 - \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \le \beta_\ell,$$
$$\text{for } \ell = 1, \dots, L. \quad (P2)$$

The feasibility of the problem in (P2) can be verified via the NP lemma or its generalization to more than one side condition [38, Sec. 3.6]. As an example, consider a binary modulation classification scenario. Let $\{\delta_{NP,1}^{(\ell)}(\cdot)\}_{\ell=1}^2$ denote the optimal NP decision rule under type-1 error constraint β_1, i.e., $\int \delta_{NP,1}^{(2)}(\boldsymbol{x}) f^{(1)}(\boldsymbol{x}) d\boldsymbol{x} = \beta_1$. If its type-2 error satisfies $\int \delta_{NP,1}^{(1)}(\boldsymbol{x}) f^{(2)}(\boldsymbol{x}) d\boldsymbol{x} \le \beta_2$, then the problem in (P2) is feasible with the constraints β_1 and β_2. By relaxing the constraints compared to the NP classifier, it is desired to improve the worst case symbol decoding performance.

We follow a similar procedure to that in Section III-A and first derive the optimal symbol decoder for fixed modulation classifier, which is presented in the following lemma.

Lemma 2: For any feasible modulation classification rule specified by $\boldsymbol{\delta}$, the optimal symbol decoders that minimize $\mathcal{J}_m(\boldsymbol{\delta}, \boldsymbol{\phi})$ are obtained as the MAP decision rules $\{\phi_{MAP}^{(\ell)}(\cdot)\}_{\ell=1}^L$ given in (7).

Proof: Since the constraint functions in (P2) do not depend on the symbol decoding rules $\{\phi^{(\ell)}(\cdot)\}_{\ell=1}^L$, the optimal symbol decoders can be derived by restricting our attention to the objective function $\mathcal{J}_m(\boldsymbol{\delta}, \boldsymbol{\phi})$ in (P2). Then, the problem of interest becomes

$$\min_{\boldsymbol{\phi}} \mathcal{J}_m(\boldsymbol{\delta}, \boldsymbol{\phi}) = \min_{\boldsymbol{\phi}} \max_{\ell \in \{1,\dots,L\}} \frac{E_\ell[C(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})\delta^{(\ell)}(\boldsymbol{x})]}{E_\ell[\delta^{(\ell)}(\boldsymbol{x})]}. \quad (21)$$

In the proof of Lemma 1, it is shown that $E_\ell[C(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})\delta^{(\ell)}(\boldsymbol{x})]$ is minimized when MAP rule based symbol decoders are employed under each modulation scheme. Furthermore, for fixed modulation classification rule, the correct classification probability of the ℓth modulation scheme, i.e., $E_\ell[\delta^{(\ell)}(\boldsymbol{x}) = 1]$, is constant. Consequently, for any set of symbol decoding rules $\{\phi^{(\ell)}(\cdot)\}_{\ell=1}^L$ and fixed modulation classification rule $\{\delta^{(\ell)}(\cdot)\}_{\ell=1}^L$, the following holds for all $\ell \in \{1, \dots, L\}$:

$$\frac{E_\ell[C(\phi^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})\delta^{(\ell)}(\boldsymbol{x})]}{E_\ell[\delta^{(\ell)}(\boldsymbol{x})]}$$
$$\ge \frac{E_\ell[C(\phi_{MAP}^{(\ell)}(\boldsymbol{x}), \boldsymbol{m})\delta^{(\ell)}(\boldsymbol{x})]}{E_\ell[\delta^{(\ell)}(\boldsymbol{x})]}, \quad (22)$$

where $\phi_{MAP}^{(\ell)}(\cdot)$ is as given in (7). Taking the maximum overall $\ell \in \{1, \dots, L\}$ on both sides, it is seen that the solution to (21) is given by the MAP rule based symbol decoders. ∎

After inserting the optimal MAP decoders into the objective function, the optimization problem (P2) is written as

$$\min_{\boldsymbol{\delta}} \max_{\ell \in \{1,\ldots,L\}} \left\{ \frac{\int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}}{\int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}} \right\}$$

$$\text{subject to } 1 - \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \beta_\ell,$$

$$\text{for } \ell = 1, \ldots, L \qquad (23)$$

where the posterior symbol decoding cost $\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x})$ is as given in (15). An optimization problem in the form of (23) is considered by Jajamovich et al. in the context of joint detection and estimation [24]. Based on the result in [24, Th. 7], the globally optimal solution to (23) is given by

$$\delta^{(j)}(\boldsymbol{x}) = 1$$

$$\text{if } j = \underset{\ell \in \{1,\ldots,L\}}{\arg \min} f^{(\ell)}(\boldsymbol{x}) \left(a_\ell^1 (\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x}) - \gamma) - a_\ell^2 \right) \quad (24)$$

and $\delta^{(j)}(\boldsymbol{x}) = 0$ otherwise, for $j = 1, \ldots, L$. In (24), $(a_\ell^1)_{\ell=1}^L$ and $(a_\ell^2)_{\ell=1}^L$ are nonnegative multipliers selected to satisfy $\sum_{\ell=1}^L a_\ell^1 + \sum_{\ell=1}^L a_\ell^2 = 1$ as well as the constraints on modulation classification error probabilities and γ denotes the optimal value of the solution to (23). The nonnegative multipliers $\boldsymbol{a} = [(a_\ell^1)_{\ell=1}^L, (a_\ell^2)_{\ell=1}^L]$ can be found along with the optimal value γ by performing a numerical search as described in [24, Sec. III.B]. To put it briefly, this can be accomplished by first discretizing the standard $(2L-1)$-simplex and then, for each point $\boldsymbol{a}$ in the discretized grid, computing the decision rules in (24) for the current value of γ, checking whether the resulting decision rules satisfy

$$\frac{\int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}}{\int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}} \leq \gamma$$

$$1 - \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \beta_\ell \qquad (25)$$

for all $\ell = 1, \ldots, L$, decreasing (increasing) the value of γ via a line or bisection search if all conditions are (not) satisfied, and repeating these steps in an iterative manner until γ can no further decreased. All in all, the modulation classification rule specified by (24) together with the MAP decoders in (7) characterize the solution to (P2). In Algorithm 1, a summary of the proposed techniques in this part and the previous part is provided.

It is noted that the optimal modulation classifier specified in (24) depends on *a posteriori* probabilities of the transmitted symbols through $\{\mathrm{C}_{\mathrm{p}}^{(\ell)}(\cdot)\}_{\ell=1}^L$ as well as the likelihood of the observed samples (i.e., $\{f^{(\ell)}(\cdot)\}_{\ell=1}^L$) similar to the previous part (see (17) and (18)). On the other hand, as opposed to the previous part where there is only one Lagrange multiplier term to be determined (based on the KKT conditions) in the modulation classification rule, in this case, multiple coefficients appear in the classification rule. Moreover, it should be remarked that the proposed modulation classifier is jointly optimal in the sense that no other modulation classifier can deliver a lower maximum symbol error probability in the case of correct modulation classification while satisfying given constraints on individual modulation classification error probabilities. Furthermore, if the individual modulation classification

error probabilities corresponding to the NP classifier (or its generalization when $L > 2$) are used as constraints in (P2), the conventional method of treating modulation classification separately from the subsequent symbol decoding is optimal. By relaxing the constraints, the symbol decoding performance quantified by the worst case metric specified in (20) can be improved.

Algorithm 1 Joint Modulation Classification and Symbol Decoding

---

procedure JOINTMODCLASSSYMDEC($\boldsymbol{x}$)
  for each candidate modulation $\ell \in \{1, 2, \ldots, L\}$ **do**
    for $k = 1, 2, \ldots, K$ **do**
      Compute $f^{(\ell)}(x_k \mid \mu_m^{(\ell)}) = f_\eta^{(\ell)}(x_k - \mu_m^{(\ell)})$
        $\forall m \in \{1, 2, \ldots, M^{(\ell)}\}$
      Compute $p^{(\ell)}(\mu_m^{(\ell)} \mid x_k)$ using (6)
        $\forall m \in \{1, 2, \ldots, M^{(\ell)}\}$
      Decode the symbol using (7), i.e.,
        $\hat{m}_k^{(\ell)} = \arg\max_{m \in \{1,\ldots,M^{(\ell)}\}} p^{(\ell)}(\mu_m^{(\ell)} \mid x_k)$
    end for
    Compute $f^{(\ell)}(\boldsymbol{x})$ using (2) and (3)
    Compute $\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x})$ by inserting previously computed
      $p^{(\ell)}(\mu_{\hat{m}_k^{(\ell)}}^{(\ell)} \mid x_k)$ values into (15)
  end for
  Select the modulation format based on (17) and (18) or
    based on (24), in the known or unknown modulation
    scheme priors cases, respectively
  Output the symbols $\{\hat{m}_k^{(\hat{\ell})}\}_{k=1}^K$ where $\hat{\ell}$ denote the index
    of selected modulation format
end procedure

C. Discussion

In this part, we present some remarks about the proposed framework. To this end, let an adaptive modulation system be configured such that the employed modulation format can be changed after every transmitted K symbols. Suppose that a sequence of length $N \times K$ symbols with $N > 1$ is observed according to this adaptive modulation scenario where the received signal sequence between indices $1 + (n-1)K$ and nK for $n \in \{1, \ldots, N\}$ follows the signal model given in (1). For convenience, each of these signal sequences of length K is termed as a single block in the following. Suppose also that the channel parameters and the noise levels are such that if the conventional technique is utilized, the average probability of symbol error in the case of correct classification is not acceptable for the application in hand. In order to reduce the symbol error to an acceptable level, the proposed approach in (P1) compromises modulation classification performance. By relaxing the constraint on the modulation classification error w.r.t. that of the MAP based classifier, the resulting modulation classifier is allowed to decide in favor of a modulation scheme that delivers a lower individual symbol error rate (even though it may not be the most probable modulation scheme for the observed block) optimally in the sense that the symbol error rate averaged over all candidate constellations is minimized subject to the relaxed constraint

TABLE I
NUMBER OF OPERATIONS REQUIRED FOR CONVENTIONAL AND PROPOSED TECHNIQUES
(FINAL EXPRESSIONS CORRESPOND TO THE NUMERICAL EXAMPLES IN SECTION VI)

	# of function evaluations	# of additions or subtractions	# of multiplications or divisions	# of maxima or minima
Conventional	$KM_\mathrm{s}= 112K$	$K(2M_\mathrm{s} - L)$ $\approx 2KM_\mathrm{s}= 224K$	$2K(M_\mathrm{max} + L) - L$ $\approx 2KM_\mathrm{max}= 128K$	$K(M_\mathrm{max} - 1)$ $\approx KM_\mathrm{max}= 64K$
Proposed	$KM_\mathrm{s}= 112K$	$2KM_\mathrm{s}= 224K$	$2K(M_\mathrm{s} + L)$ $\approx 2KM_\mathrm{s}= 224K$	$K(M_\mathrm{s} - L)$ $\approx KM_\mathrm{s}= 112K$

on the modulation classification error. (Recall that since all of the decoded symbols in that block will be erroneous if the modulation scheme is not identified correctly and the constellation points of candidate modulation schemes do not overlap, it is necessary to constrain modulation classification errors.) Another way to think of this is as follows: Suppose that a block of signal samples that falls around a small neighborhood of the decision boundary of the MAP based classifier is observed at the receiver. As far as the MAP based modulation classifier is concerned, this block can be assigned to any modulation scheme on either side of the decision boundary since that would not alter the performance metric of the MAP based classifier noticeably. On the other hand, one of these modulation schemes is likely to yield a lower symbol error probability than the other in the case of a correct decision on the employed modulation scheme. In that respect, the overall symbol error performance in the case of correct classification can be reduced by assigning the received blocks close to the decision boundary to the modulation scheme with smaller symbol error score while controlling the erroneous symbol decoding due to an incorrect classification decision with the constraint on the tolerable modulation classification error.

IV. COMPUTATIONAL COMPLEXITY

In this section, we examine the computational complexity of joint modulation classification and symbol decoding. In order to implement the proposed methods, it is required to find λ or the pair $\boldsymbol{a}$ and γ when the prior probabilities of the modulation formats are known or unknown, respectively. In practice, these parameters can be determined in an offline manner for possible system configurations (e.g., different SNR and K values) and the corresponding receiver can directly use these values. Therefore, in the following, we assume that these parameters have been computed and focus on the computational complexity of the joint modulation classification and symbol decoding scheme. Then, the complexity of computing these parameters is explained.

In order to identify the employed modulation scheme, both of the proposed techniques need to compute the likelihood scores (i.e., $\{f^{(\ell)}(\cdot)\}_{\ell=1}^L$) and the posterior symbol decoding costs (i.e., $\{\mathrm{C}_\mathrm{P}^{(\ell)}(\cdot)\}_{\ell=1}^L$) based on the observed samples for each candidate modulation format. Since $\mathrm{C}_\mathrm{P}^{(\ell)}(\cdot)$ depends on the *a posteriori* probabilities of the symbols selected by the MAP decoding rule corresponding to the ℓth modulation scheme, the received signal sequence is first decoded using the MAP decoding rule of the ℓth modulation scheme and then *a posteriori* probability values achieved by the decoded symbols are inserted into (15). With this approach,

modulation classification and symbol decoding stages are combined. Thus, the overall computational complexity of the proposed techniques depends on the number of operations required to obtain $\{f^{(\ell)}(\cdot)\}_{\ell=1}^L$ and $\{\mathrm{C}_\mathrm{P}^{(\ell)}(\cdot)\}_{\ell=1}^L$. They are given in Table I, where function evaluations represent the computation of $f_\eta^{(\ell)}(\cdot)$, and M_s and M_max are defined as $M_\mathrm{s} \triangleq \sum_{\ell=1}^L M^{(\ell)}$ and $M_\mathrm{max} \triangleq \max_{\ell \in \{1,\dots,L\}} M^{(\ell)}$, respectively. In the table, the particular values corresponding to the numerical examples in Section VI are also presented.

On the other hand, the conventional techniques decide on the modulation format based only on the likelihood of the observed samples and then decode the symbols according to the MAP symbol detector of the selected modulation format. Therefore, the main difference between the conventional and the proposed techniques in terms of complexity arise from the fact that the conventional techniques need to perform symbol decoding only under the selected modulation scheme whereas the proposed techniques perform symbol decoding under each candidate to obtain $\{\mathrm{C}_\mathrm{P}^{(\ell)}(\cdot)\}_{\ell=1}^L$. Yet, since the MAP symbol decoders under each modulation format needs to be implemented in practice, the complexity of the hardware for the proposed techniques can be reduced by sharing the resources for modulation classification and symbol decoding stages.

From Table I, it is seen that the number of samples (K) and the sum of modulation orders (M_s) determine the complexity of the proposed techniques. On the other hand, as the conventional techniques decode symbols only under the selected modulation format, the total number of multiplication and maximum operations depend on M_max instead of M_s considering the worst case scenario when the highest order modulation format is selected.

Next, we examine the computational complexity of the procedure for setting the parameters that appear in the modulation classification rules, i.e., λ, $\boldsymbol{a}$ and γ. For (P1), Monte Carlo simulations can be performed to determine λ. Towards this goal, we consider the discretized interval $[0, \lambda_\mathrm{max}]$ where λ_max is such that $\mathcal{H}_\mathrm{a}(\boldsymbol{\delta}(\lambda_\mathrm{max})) \approx \alpha_\mathrm{B}$. For a given λ in this discretized interval, we compute the decision rule for each Monte Carlo trial and then empirically obtain the average probability of modulation classification error using the results of these trials. Next, we repeat this procedure for each λ in this discretized interval until $|\mathcal{H}_\mathrm{a}(\boldsymbol{\delta}(\lambda)) - \alpha| \leq \epsilon$ for some small ϵ. Treating each iteration as $\mathcal{O}(1)$, the computational complexity of this complete procedure is given by $\mathcal{O}(N_T N_G^\lambda)$ where N_T and N_G^λ denote the number of Monte Carlo trials and the size of the discretized grid, respectively. For (P2), the optimal values of $\boldsymbol{a}$ and γ are determined based on the algorithm

in [24]. In this algorithm, a single iteration of the bisection search requires $\mathcal{O}(N_T N_G^{\boldsymbol{a}})$ computations where $N_G^{\boldsymbol{a}}$ denotes the size of the discretized grid for $\boldsymbol{a}$ and a few iterations is in practice sufficient for convergence. It should be noted that the complexity of this procedure can be significantly higher than the previous case since the discretization for $\boldsymbol{a}$ is on a standard $(2L-1)$-simplex whereas the discretization for λ is simply over an interval on the real line.

V. EXTENSION

In this section, an extension to the proposed framework is considered by utilizing a different cost function to evaluate the symbol decoding performance. In Section III-A and Section III-B, the symbol decoding performance of the proposed techniques is measured by the cost function defined in (8), which appears in the objective functions of (P1) and (P2). The cost function in (8) essentially counts the number of symbol errors made by the employed detector out of K symbol transmissions that constitute the received signal sequence in (1). In other words, the resulting performance metric penalizes each erroneous symbol decoding separately in a given block of length K. However, in certain applications, rather than individual symbol errors, it may be important to decode the whole block of symbols (e.g., codewords, frames, etc.) correctly. In order to meet the requirements of such applications, the following cost function is introduced:

$$\mathrm{C_b}(\boldsymbol{\phi}^{(\ell)}(\boldsymbol{x}), \boldsymbol{m}) = \mathbb{1}_{\{\boldsymbol{\phi}^{(\ell)}(\boldsymbol{x}) \neq \boldsymbol{m}\}} \qquad (26)$$

which yields 1 if the transmitted and decoded messages do not agree and 0 otherwise, where $\boldsymbol{\phi}^{(\ell)}(\boldsymbol{x}) \triangleq [\phi^{(\ell)}(x_1), \dots, \phi^{(\ell)}(x_K)]^T$.

In order to characterize the solution to (P1) and (P2) when the objective function is computed using (26) instead of (8), we begin by deriving the optimal symbol decoding rule for a fixed modulation classifier. To that end, the optimality of the MAP decoding is shown first in the following. Furthermore, it is sufficient to only consider (P1) here since the optimality in the case of (P2) is clear following a similar reasoning as in the proof of Lemma 2.

For a fixed modulation classification rule $\boldsymbol{\delta}$, the optimization problem in (P1) with the new cost function is written as

$$\min_{\phi^{(\ell)}} \int \delta^{(\ell)}(\boldsymbol{x}) \sum_{m_1=1}^{M^{(\ell)}} \cdots \sum_{m_K=1}^{M^{(\ell)}} \mathbb{1}_{\{\boldsymbol{\phi}^{(\ell)}(\boldsymbol{x}) \neq \boldsymbol{m}\}}$$
$$\left(\prod_{k=1}^{K} \frac{1}{M^{(\ell)}} f^{(\ell)}(x_k \mid \mu_{m_k}^{(\ell)}) \right) d\boldsymbol{x}. \qquad (27)$$

After some manipulation, we get

$$\max_{\phi^{(\ell)}} \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x})$$
$$\left(\sum_{m_1=1}^{M^{(\ell)}} \mathbb{1}_{\{\phi^{(\ell)}(x_1)=m_1\}} p^{(\ell)}(\mu_{m_1}^{(\ell)} \mid x_1) \right)$$
$$\vdots$$
$$\left(\sum_{m_K=1}^{M^{(\ell)}} \mathbb{1}_{\{\phi^{(\ell)}(x_K)=m_K\}} p^{(\ell)}(\mu_{m_K}^{(\ell)} \mid x_K) \right) d\boldsymbol{x}. \qquad (28)$$

It is now clear from (28) that the optimal symbol decoder is given by the MAP rule, i.e., $\phi_{\mathrm{MAP}}^{(\ell)}(\cdot)$.

Having identified the symbol decoder, we proceed with the derivation of the modulation classification rule. By inserting MAP based symbol decoders into the objective function (27), the problem reduces to

$$\min_{\boldsymbol{\delta}} \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x}$$
$$\text{subject to } 1 - \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \alpha \qquad (29)$$

where $\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x})$ denotes the posterior cost associated with the optimal MAP based symbol decoder for the ℓth constellation computed using the cost function in (26) and it is given by

$$\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x}) = 1 - \prod_{k=1}^{K} \max_{m \in \{1, \dots, M^{(\ell)}\}} p^{(\ell)}(\mu_m^{(\ell)} \mid x_k). \qquad (30)$$

Notice that the problem in (29) is the same as that in (16) after replacing $\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x})$ with $\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x})$. Therefore, the solution to (29) is given by the modulation classification rule specified in (17) and (18) after substituting $\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x})$ in place of $\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x})$. Furthermore, a similar approach can be employed in the case of (P2) leading to the modulation classification rule specified by (24), where $\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x})$ is again replaced with $\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x})$.

Moreover, we can analyze an interesting case in which the optimization problem does not involve any constraints. Recall that in (P1), the objective function measures the symbol error probability in the case of correct modulation classification while the constraint ensures a given level of modulation classification accuracy. Instead of such a constrained problem, an unconstrained problem can also be formulated by combining these two performance measures into a single objective function with equal weights. More specifically, consider the following optimization problem

$$\min_{\boldsymbol{\delta}, \boldsymbol{\phi}} \mathcal{J}_{\mathrm{a,b}}(\boldsymbol{\delta}, \boldsymbol{\phi}) + \mathcal{H}_{\mathrm{a}}(\boldsymbol{\delta}) \qquad (31)$$

where

$$\mathcal{J}_{\mathrm{a,b}}(\boldsymbol{\delta}, \boldsymbol{\phi}) \triangleq \sum_{\ell=1}^{L} c_\ell \mathrm{E}_\ell [\mathrm{C}_{\mathrm{b}}(\boldsymbol{\phi}^{(\ell)}(\boldsymbol{x}), \boldsymbol{m}) \delta^{(\ell)}(\boldsymbol{x})] \qquad (32)$$

and $\mathcal{H}_{\mathrm{a}}(\boldsymbol{\delta})$ is as given in (10). Under the assumption that the modulation classification errors automatically lead to symbol errors (i.e., when constellation points of candidate modulation schemes do not overlap), the solution of the optimization problem in (31) can be viewed as a joint mechanism that minimizes the total probability of symbol error composed of those that occur when modulation scheme is correctly identified and those that occur when a modulation classification error is made.

Noticing that the second term in (31) does not depend on $\boldsymbol{\phi}$, and based on the previous discussions (see (28)), it can be concluded that the MAP rule, i.e., $\phi_{\mathrm{MAP}}^{(\ell)}(\cdot)$, solves the optimization problem in (31) for a fixed modulation classification rule $\boldsymbol{\delta}$. Next, after inserting the optimal MAP decoders,

the optimization problem in (31) reduces to

$$\min_{\boldsymbol{\delta}} \ \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) \left(\mathrm{C}_{\mathrm{p,b}}^{(\ell)}(\boldsymbol{x}) - 1 \right) d\boldsymbol{x}. \quad (33)$$

From (33), it can be seen that the optimal modulation classification rule is given by

$$\delta^{(j)}(\boldsymbol{x}) = 1$$
$$\text{if } j = \operatorname*{arg\,min}_{\ell \in \{1,\dots,L\}} c_\ell f^{(\ell)}(\boldsymbol{x}) \left(\mathrm{C}_{\mathrm{p}}^{(\ell)}(\boldsymbol{x}) - 1 \right) \quad (34)$$

and $\delta^{(j)}(\boldsymbol{x}) = 0$ otherwise, for $j = 1, \dots, L$.

Note that the optimal modulation classifier in the constrained problem contains a Lagrange multiplier term which essentially determines the tradeoff between the symbol decoding performance in the case of correct modulation classification and the average probability of misclassification. On the other hand, it is observed that the optimal modulation classifier derived in (34) is a special case of the constrained problem when $\lambda = 1$, which means that it does not permit tradeoffs between the considered performance measures of symbol decoding and modulation classification.

In the following, we provide another perspective to the unconstrained optimization problem in (31). Combining both terms in the objective function and after some manipulation, the optimization problem in (31) can be expressed as

$$\max_{\boldsymbol{\delta},\boldsymbol{\phi}} \sum_{\ell=1}^{L} c_\ell \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x})$$
$$\left(\sum_{m_1=1}^{M^{(\ell)}} \mathbb{1}_{\{\phi^{(\ell)}(x_1)=m_1\}} p^{(\ell)}(\mu_{m_1}^{(\ell)} \mid x_1) \right)$$
$$\vdots$$
$$\left(\sum_{m_K=1}^{M^{(\ell)}} \mathbb{1}_{\{\phi^{(\ell)}(x_K)=m_K\}} p^{(\ell)}(\mu_{m_K}^{(\ell)} \mid x_K) \right) d\boldsymbol{x}. \quad (35)$$

Furthermore, the optimization problem in (35) can alternatively be written as

$$\max_{\ell \in \{1,\dots,L\}} \ \max_{m_k \in \mathcal{M}^{(\ell)}, \, k=1,\dots,K} c_\ell f^{(\ell)}(\boldsymbol{x}) \prod_{k=1}^{K} p^{(\ell)}(\mu_{m_k}^{(\ell)} \mid x_k)$$
$$= \max_{\ell \in \{1,\dots,L\}} \ \max_{m_k \in \mathcal{M}^{(\ell)}, \, k=1,\dots,K} c_\ell \prod_{k=1}^{K} \frac{1}{M^{(\ell)}} f^{(\ell)}(x_k \mid \mu_{m_k}^{(\ell)})$$
$$= \max_{\ell \in \{1,\dots,L\}} \ \max_{m_k \in \mathcal{M}^{(\ell)}, \, k=1,\dots,K}$$
$$p(\mathsf{D} = \mathcal{S}^{(\ell)}, \mu_{m_1}^{(\ell)}, \dots, \mu_{m_K}^{(\ell)}, \boldsymbol{x})$$
$$= \max_{\ell \in \{1,\dots,L\}} \ \max_{m_k \in \mathcal{M}^{(\ell)}, \, k=1,\dots,K}$$
$$p(\mathsf{D} = \mathcal{S}^{(\ell)}, \mu_{m_1}^{(\ell)}, \dots, \mu_{m_K}^{(\ell)} \mid \boldsymbol{x}) \quad (36)$$

where $\mathcal{M}^{(\ell)} \triangleq \{1, \dots, M^{(\ell)}\}$. The final expression in (36) states that the optimal modulation classification and symbol decoding rules need to be such that the joint *a posteriori* probability of modulation format and symbol sequence are

maximized given the observed signal sequence. Due to the fact that a maximization over the joint *a posteriori* probability is performed, this technique is referred to as *joint MAP*. All in all, using the cost function (26) in the objective function of (P1) and taking $\lambda = 1$ in the modulation classification rule, the *joint MAP* technique is obtained as a special case.

VI. Numerical Results

In this section, numerical examples are presented to examine the performance of the proposed solutions to the joint modulation classification and symbol decoding problem. In particular, to illustrate the effectiveness of the proposed techniques, performance comparisons are conducted against the conventional techniques. Note that both the proposed techniques and the conventional techniques decode the symbols based on the MAP decoder for the recovered modulation scheme. The difference between the formulations is related to the modulation classification rules.[5] Therefore, in the following, the proposed techniques in Section III-A and Section III-B are compared against the conventional techniques that classify the modulation scheme based on the Bayes modulation classifier and the minimax[6] modulation classifier, respectively, which are described in Section II-B.

The setup considered in the simulations is as follows: There exist three possible candidate modulation schemes which are 16-QAM, 32-QAM and 64-QAM, i.e., $M^{(1)} = 16$, $M^{(2)} = 32$ and $M^{(3)} = 64$ [9], [14], [18]. The modulation schemes are assumed to have unit average power.[7] The noise samples η_k are taken as complex circular Gaussian random variables with variance $2\sigma_\ell^2$ under the ℓth modulation scheme leading to an additive white Gaussian noise (AWGN) channel. Under this setting, the SNR for the ℓth modulation scheme becomes $\mathrm{SNR}_\ell = 1/(2\sigma_\ell^2)$. The aim of allowing for different noise levels under each modulation scheme is to model the behavior of adaptive modulation systems which select the modulation format to be employed according to the channel conditions.

A. Performance in the Case of Known Priors for the Modulation Schemes

We first investigate the case in which the prior probabilities of the modulation schemes are known by solving (P1) for the previously described setup. In particular, under each modulation scheme, SNR values are set to 18 dB, and each modulation scheme is equally likely, i.e., $c_1 = c_2 = c_3 = 1/3$. For this scenario, Fig. 1a plots the average symbol decoding performance in the case of correct modulation classification with respect to the constraint on the average probability of misclassification for the proposed technique for $K = 1, 3, 10$,

---

[5] It is important to emphasize that although the resulting joint mechanisms of the proposed techniques lead to a change only in the modulation classifier compared to the conventional techniques, the proposed techniques are jointly optimal considering the introduced performance measures of both subproblems.

[6] In the numerical examples, since equal constraints on the individual probability of classification errors are assumed, i.e., $1 - \int \delta^{(\ell)}(\boldsymbol{x}) f^{(\ell)}(\boldsymbol{x}) d\boldsymbol{x} \leq \beta$ for all $\ell \in \{1, \dots, L\}$, the minimum of such β results in minimax decision rule [20].

[7] Different but known power levels can also be considered as there are no assumptions about the power levels in Section III.
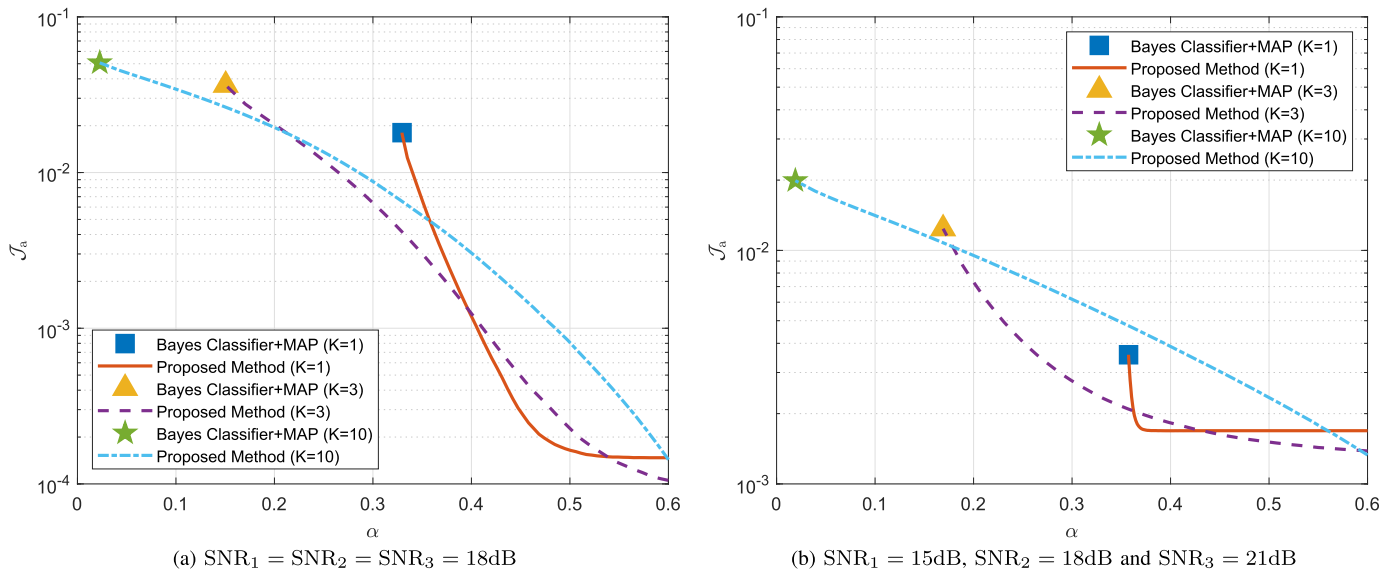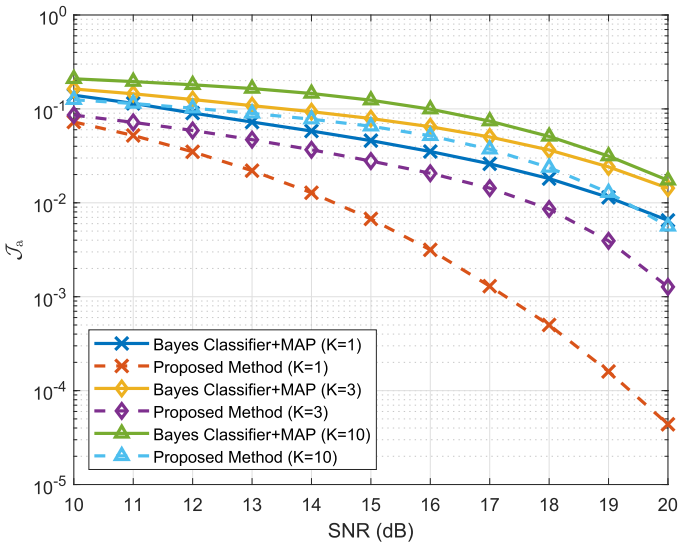
Fig. 1.   Comparison of the conventional and the proposed approaches in the case of known modulation scheme priors.

together with the corresponding performances of the conventional technique, which are represented by the single points (as they yield a fixed performance for each K). It is observed that for each value of K, the symbol decoding performance can be improved by relaxing the constraint on the probability of modulation classification error. Therefore, it is verified that a tradeoff between the subproblems becomes possible via the proposed approach. Moreover, as the proposed technique is jointly optimal, no other method can operate below the plotted curves. Therefore, these plots provide performance benchmarks for any other joint mechanism. Furthermore, it is observed that when $K = 1$, a small relaxation (e.g., changing $\alpha = \alpha_{\mathrm{B}}$ with $\alpha = 0.5$) in the constraint compared to the Bayesian classifier leads to a significant improvement in the symbol decoding performance. On the other hand, when further relaxation (e.g., changing $\alpha = 0.5$ with $\alpha = 0.6$) is applied, no significant improvement on the symbol decoding performance is obtained. On the other hand, when K is increased, the relaxation required for a certain order of magnitude reduction in the symbol decoding performance measure becomes larger. Therefore, if a significant improvement on the symbol decoding performance is desired in the case of a large K, the constraint should not be close to α_{B}. In fact, as can be observed from the plot, by taking the constraint as large as 0.5, it is possible to reduce the symbol error probability in the case of correct classification to a value lower than 10^{-3}, which is considered to be an acceptable level for a variety of applications, for $K = 10$.

In Fig. 1b, we change the SNR values to $\mathrm{SNR}_1 = 15\,\mathrm{dB}$, $\mathrm{SNR}_2 = 18\,\mathrm{dB}$ and $\mathrm{SNR}_3 = 21\,\mathrm{dB}$ and plot the average symbol decoding performance when the modulation format is correctly identified versus the constraint on the probability of misclassification for $K = 1, 3, 10$. The reason for considering a different set of SNR values is to show the effectiveness of the proposed technique under a variety of conditions. It should also be emphasized that under the modulation format with

a higher order, the channel is assumed to be less noisy (i.e., higher SNR), which is consistent with the operation of adaptive modulation systems. The first observation from Fig. 1b is that by relaxing the constraint on the probability of misclassification, an improvement on the symbol decoding performance can be obtained compared to the conventional technique for different values of K. In other words, the tradeoff between the subproblems is enabled under a variety of conditions (i.e., for different SNR and K values). Similar to the previous case, for $K = 1$, a slight relaxation is sufficient to set the symbol decoding risk to a value that is close to the lowest possible value. On the other hand, for a large K, it is required to apply a further relaxation in order to reduce the symbol decoding risk, for instance to a value close to 10^{-3}. Moreover, the relaxation in the constraint in this case is not as effective as the relaxation in the previous case, especially for low K's. The intuition behind this observation can be explained as follows: In the previous case, since SNR values are the same, the symbol decoding performance under 64-QAM is worse than that under 16-QAM and 32-QAM. Hence, the overall symbol decoding performance is mainly dominated by the performance of 64-QAM. This provides enhanced flexibility to reduce the overall symbol decoding risk by making more classification errors under 64-QAM. On the other hand, there is no such flexibility in the second case due to the fact that each modulation scheme has comparable symbol decoding performance.

We next investigate the performance of the proposed technique together with the conventional technique for a range of SNR values in which each modulation scheme operates under the same SNR condition. For a given SNR value, the conventional technique yields a fixed probability of modulation classification error, i.e., α_{B}. In order to obtain a tradeoff via the proposed technique, a fixed percentage of correct modulation format decisions made by the conventional technique is allowed to be incorrect in the following simulations.

(a) Symbol decoding performance versus SNR
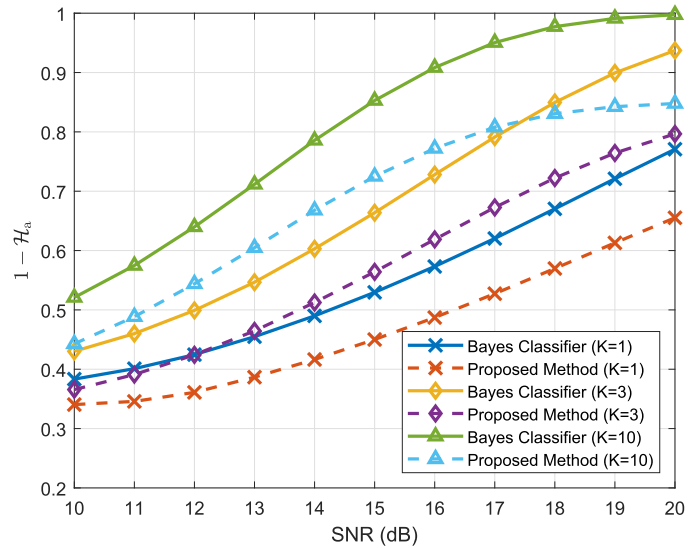
(b) Average probability of correct modulation classification versus SNR

Fig. 2. Comparison of the conventional and the proposed approaches in the case of known modulation scheme priors for $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = \mathrm{SNR}$ and $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$.

In particular, we set α as $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$; hence, the constraint in (P1) becomes $(1 - \mathcal{H}_{\mathrm{a}}(\boldsymbol{\delta})) \geq 0.85(1 - \alpha_{\mathrm{B}})$. Under this scenario, Fig. 2a plots symbol decoding performance in the case of correct classification versus SNR and Fig. 2b plots probability of correct modulation classification versus SNR for each technique and for $K = 1, 3, 10$. From Fig. 2a, it can be observed that at low SNR values both techniques deliver similar symbol error rate, and thus relaxation is not very effective. On the other hand, in the high SNR regime, relaxation in the constraint becomes more effective, especially for small K values. Furthermore, an interesting observation is that symbol decoding performance, measured by $\mathcal{J}_{\mathrm{a}}$ in the case of correct classification, degrades for the conventional method as the number of samples K increases. This can be explained by the fact that the modulation classification performance of the conventional method improves as K increases, and consequently, noisy samples that are prone to erroneous symbol decoding are included in the computation of $\mathcal{J}_{\mathrm{a}}$.

In order to illustrate the effect of number of samples K, Fig. 3 plots the average symbol decoding performance in the case of correct classification with respect to K for $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$. It is noted that the most significant improvement in the symbol decoding performance with the proposed technique occurs when K is small and the SNR level is high. In addition, it is seen that $\mathcal{J}_{\mathrm{a}}$ becomes steady with respect to K when K is large. This is due to the fact that the modulation classification performance of the conventional method remains the same for large values of K, and consequently, the constraint on the probability of modulation classification error for the proposed method is kept the same.

B. Performance in the Case of Unknown Priors for the Modulation Schemes

We next investigate the case in which prior probabilities of the modulation schemes are unknown via solving (P2). First, the same SNR condition (i.e., $\mathrm{SNR}_1 = \mathrm{SNR}_2 =$
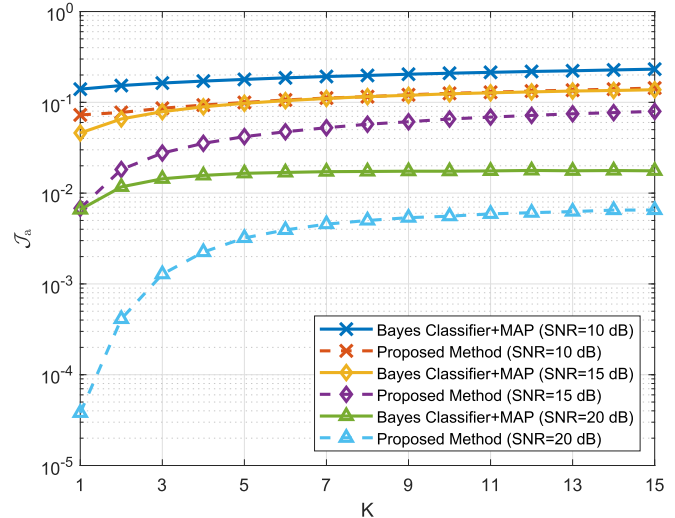

Fig. 3. Symbol decoding performance versus K for the conventional and the proposed approaches in the case of known modulation scheme priors for $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$.

$\mathrm{SNR}_3 = 18\,\mathrm{dB}$) is examined. In Fig. 4a, we plot the worst case symbol decoding performance in the case of correct classification with respect to β where $\beta_1 = \beta_2 = \beta_3 = \beta$ for the proposed technique for $K = 1, 3, 10$, together with the performance of the conventional technique. It is observed that by increasing β, significant improvements on the symbol decoding performance can be obtained for $K = 1$. For $K = 3$ and $K = 10$ on the contrary, the improvements on the symbol decoding performance with respect to β become less significant compared to $K = 1$ case. Next, SNR values are set to $\mathrm{SNR}_1 = 15\mathrm{dB}$, $\mathrm{SNR}_2 = 18\mathrm{dB}$, and $\mathrm{SNR}_3 = 21\mathrm{dB}$. In Fig. 4b, we plot the worst case symbol error probability when the modulation type is correctly identified with respect to β. It is seen that for $K = 1$, a slight relaxation in β improves the symbol decoding performance whereas further relaxation does not lead to any improvement. On the other hand, for $K = 3, 10$, multiple trends in the behavior of symbol

(a) $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = 18\mathrm{dB}$

(b) $\mathrm{SNR}_1 = 15\mathrm{dB}$, $\mathrm{SNR}_2 = 18\mathrm{dB}$ and $\mathrm{SNR}_3 = 21\mathrm{dB}$
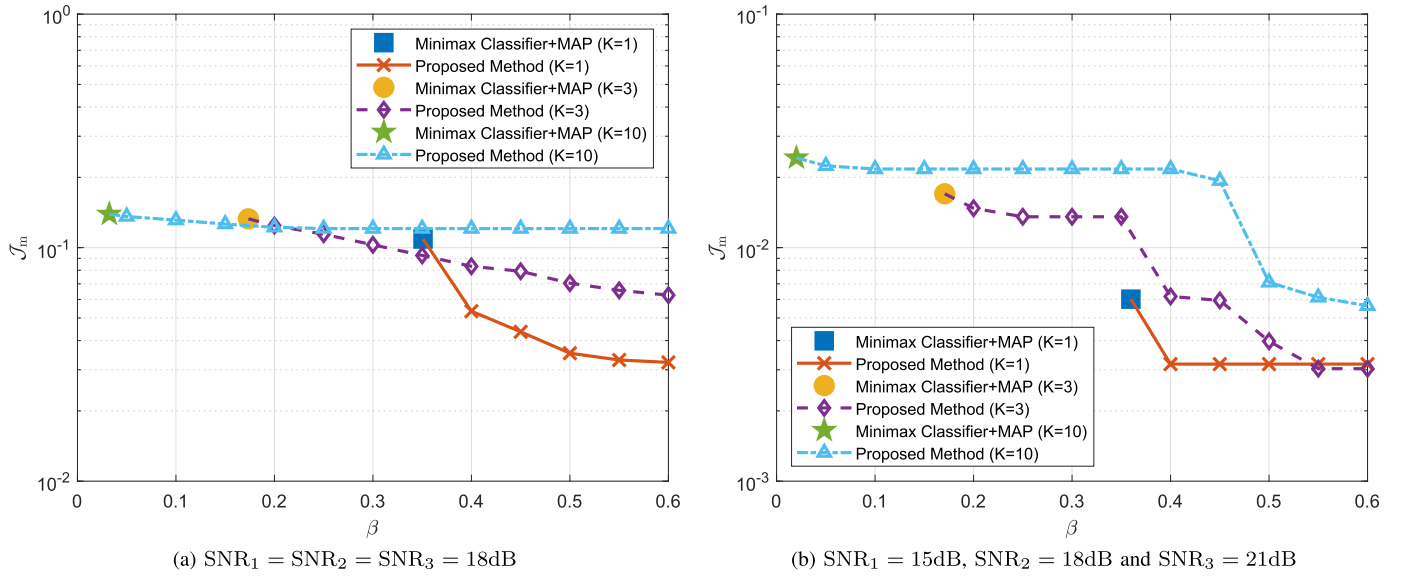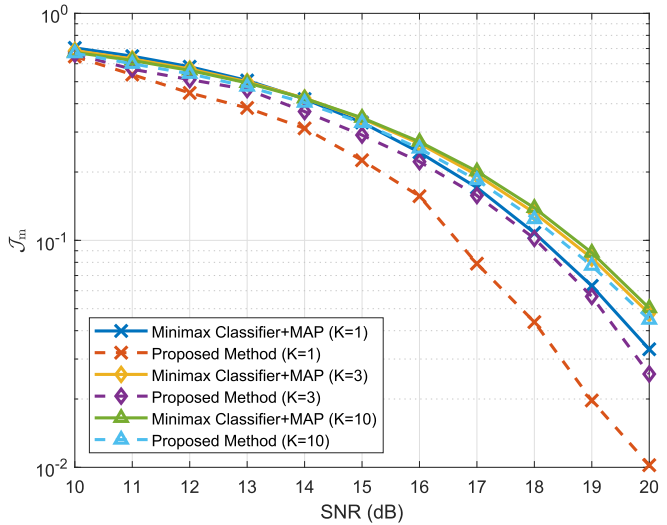
Fig. 4. Comparison of the conventional and the proposed approaches in the case of unknown modulation scheme priors.

Fig. 5. Symbol decoding performance versus SNR for the conventional and the proposed approaches in the case of unknown modulation scheme priors for $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = \mathrm{SNR}$ and $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$.
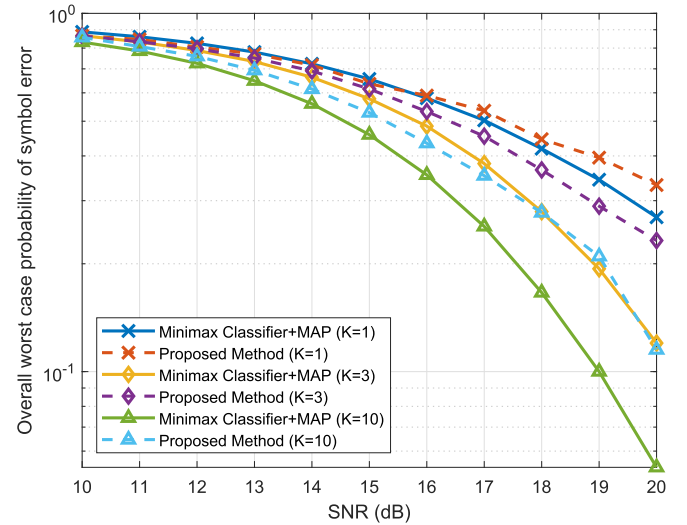
Fig. 6. Overall symbol decoding performance versus SNR for the conventional and the proposed approaches in the case of unknown modulation scheme priors for $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = \mathrm{SNR}$ and $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$.

decoding performance with respect to β are observed. This can intuitively be explained as follows: In this case, SNR levels are such that each modulation scheme has comparable symbol decoding performance. This makes it likely that the modulation format index that achieves the worst case symbol decoding performance, i.e., $\mathcal{J}_{\mathrm{m}}$, changes while relaxing the constraints, which consequently yields multiple trends in the plot of $\mathcal{J}_{\mathrm{m}}$ with respect to β.

In Fig. 5, the worst case symbol decoding performance in the case of correct modulation classification for the conventional and the proposed techniques is illustrated by varying the SNR value under which each modulation format operates. For the proposed technique, all constraints are relaxed compared to the conventional technique as follows: $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$ where β_{mm} denotes the smallest β delivered by the minimax modulation classification rule. It is seen that for different K values, the conventional technique delivers close to worst case symbol error rate in the

case of correct classification. On the other hand, an improvement on symbol decoding performance can be obtained with the proposed technique especially for small K and at high SNR values. Moreover, as SNR is increased, all individual probability of modulation classification error terms improve for the conventional technique since the conventional technique is optimized for modulation classification performance only. On the other hand, for the proposed technique, there is no such monotone behavior as the objective is to optimize for the symbol detection performance.

In Fig. 6, we plot the overall worst case symbol decoding performance, in which a modulation classification error is regarded as a symbol error, versus SNR for $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$. This overall performance metric, which is different from $\mathcal{J}_{\mathrm{m}}$ in (P2), is computed as follows: We first calculate the conditional symbol error probabilities given that the ℓth modulation scheme is employed and then take the max-

imum over ℓ to account for the worst case scenario. It should be emphasized that as the proposed technique optimizes the symbol error rate when the modulation format is correctly identified, we do not expect the proposed method to deliver the best performance in Fig. 6. In fact, as the proposed technique performs a tradeoff between the modulation classification performance and the symbol decoding performance in the case of correct classification, its overall symbol error performance deteriorates as a result of the degraded classification performance. However, the proposed technique reduces the symbol error rate in the case of correct classification measured by $\mathcal{J}_{\mathrm{m}}$ as depicted in Fig. 5. In order to explain this phenomenon further, consider a particular scenario in which the SNR is $18\,\mathrm{dB}$ and $K = 3$. For the conventional technique, the conditional symbol error rates are 8.10×10^{-5}, 1.09×10^{-2} and 1.32×10^{-1} given that 16-QAM, 32-QAM and 64-QAM, respectively, are identified correctly. On the other hand, the corresponding symbol error rates for the proposed technique become 3.25×10^{-4}, 7.12×10^{-3} and 1.02×10^{-1} considering 16-QAM, 32-QAM and 64-QAM, respectively. Therefore, the improvement under QAM-64 yields reduced $\mathcal{J}_{\mathrm{m}}$ with the proposed technique. However, the classification performance of the proposed technique is degraded, which consequently yields worse overall symbol decoding performance compared to the conventional technique.

Next, Fig. 7 plots the worst case symbol decoding performance when the modulation is correctly identified versus the number of samples for $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$. It is seen that the gap between the performances of the proposed and conventional techniques diminishes as the number of samples increases. Therefore, for the case of a high number of samples, it is preferable to employ the conventional technique as it guarantees to yield the best modulation classification performance according to the minimax criterion and at the same time its symbol decoding performance is close to the proposed technique. The proposed technique on the other hand is promising when the number of samples is small and the SNR level is high.

C. Performance in the Presence of Channel Estimation Errors

In (1), it is assumed that preprocessing tasks such as channel estimation and equalization are already performed. In case these preprocessing operations are not reliable, performance of the proposed techniques degrades. In this part, the effect of channel estimation errors is illustrated. Towards this goal, we assume that the proposed techniques are designed based on the signal model in (1), i.e., the receiver presumes that there is no channel estimation error, while the received signal is given by

$$x_k = hs_k + \eta_k, \quad k = 1, \dots, K. \tag{37}$$

In (37), h is used to model channel estimation errors and it is a randomly generated complex coefficient with amplitude $(1 + \epsilon)$ and phase κ where ϵ and κ are uniformly distributed in the interval $[-b, b]$. With this model, channel estimation errors become more significant as b increases. Moreover,
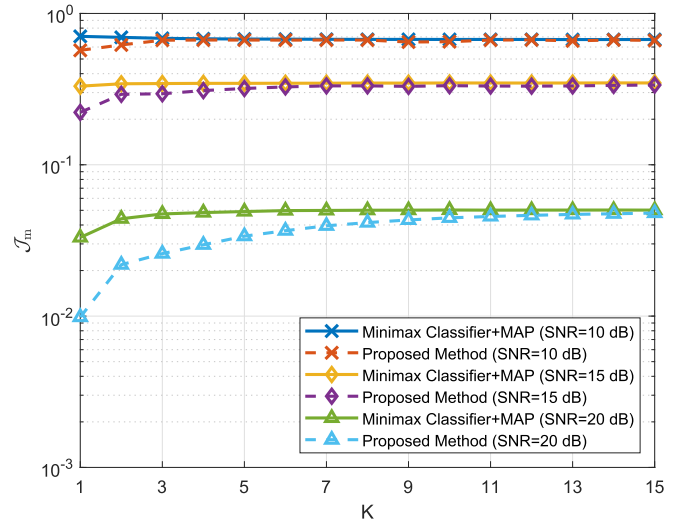

Fig. 7. Symbol decoding performance versus K for the conventional and the proposed approaches in the case of unknown modulation scheme priors for $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$.

a block fading scenario is considered in this model where the channel estimation and equalization are performed once for the complete received signal sequence.

In Fig. 8a, symbol decoding error in the case of correct classification and average probability of modulation classification error are plotted versus b for the case when the prior probabilities of the modulation schemes are known, and for $K = 3$, $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = 18\,\mathrm{dB}$ and $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$. It is observed that small channel estimation errors do not cause significant degradation for both symbol decoding and modulation classification performance. However, when the channel estimation errors become more significant, both performance measures are affected significantly. In Fig. 8b, the effect of channel estimation errors on the performance of the proposed technique in Section III-B is illustrated where K and SNR values are as before and $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$. It is seen that, similar to the previous case, small channel estimation errors do not cause significant degradation whereas large channel estimation errors deteriorate each performance measure.

D. Performance Considering the Cost Function in the Extension

In this part, performance of the conventional and the proposed techniques is presented when the cost function to measure symbol error rate is given by (26). More specifically, Fig. 9a and Fig. 9b consider the formulation in Section III-A and Section III-B, respectively, when the objective functions are as in (26) instead of (8). As expected, block symbol error performance is poorer than average symbol error performance considered in the previous examples. From Fig. 9a and Fig. 9b, it is verified that the proposed techniques corresponding to this cost function yield improved symbol decoding performance in the case of correct classification by relaxing the constraint(s) on average (or individual) probability of classification error(s) compared to the conventional techniques.
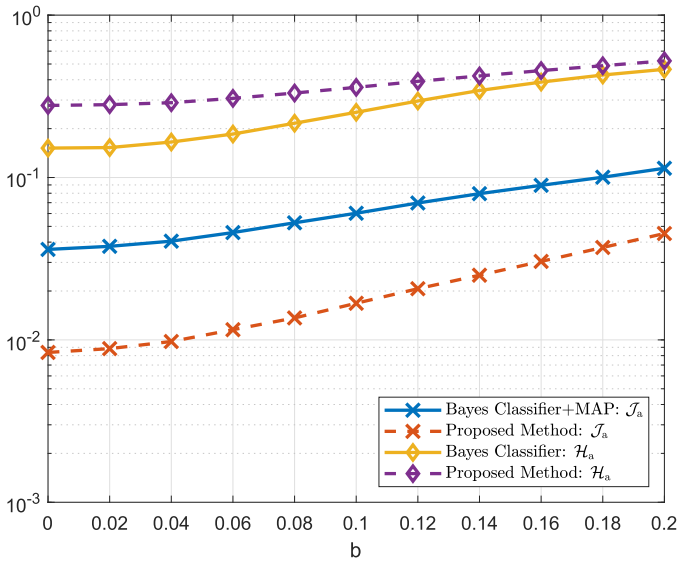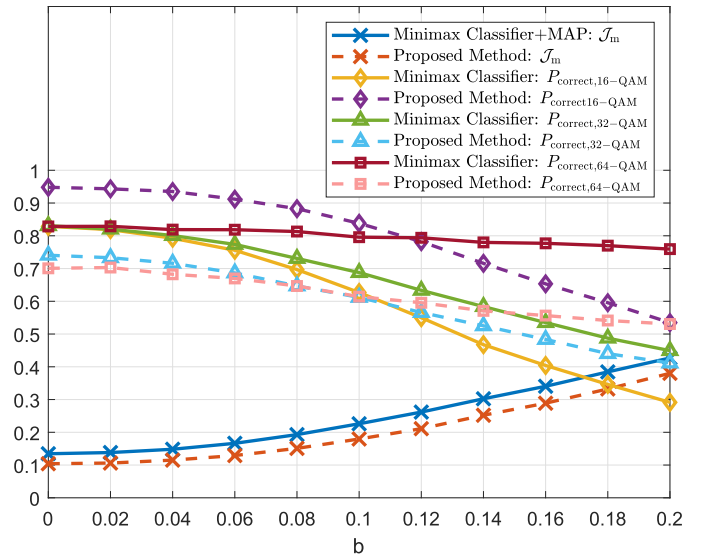
(a) Known priors for the modulation schemes for $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$

(b) Unknown priors for the modulation schemes for $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$

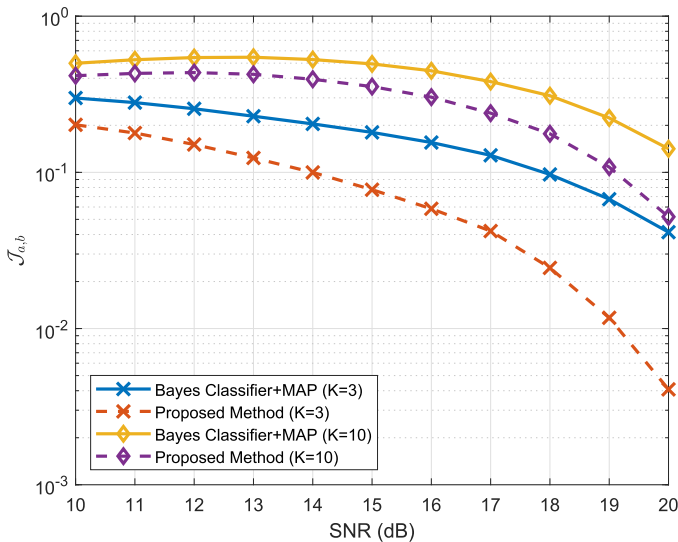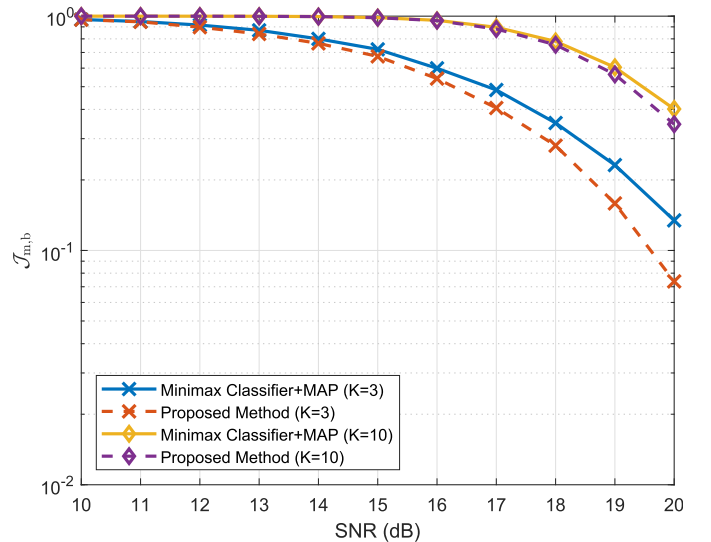Fig. 8. Performance of the conventional and the proposed approaches versus channel estimation error parameter b for $K = 3$ and $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_3 = 18\,\mathrm{dB}$.

(a) Known priors for the modulation schemes for $\alpha = 1 - 0.85(1 - \alpha_{\mathrm{B}})$

(b) Unknown priors for the modulation schemes for $\beta_1 = \beta_2 = \beta_3 = 1 - 0.85(1 - \beta_{\mathrm{mm}})$

Fig. 9. Performance of the conventional and the proposed approaches considering the cost function in (26) versus SNR.

VII. CONCLUSION

In this paper, jointly optimal approaches for solving modulation classification and symbol decoding problems have been proposed. In particular, the scenarios with both known and unknown prior probabilities of modulation schemes have been addressed to provide a generic framework. In both cases, the introduced symbol decoding performance measure has been optimized under the constraint(s) on the modulation classification performance. In this way, a tradeoff between modulation classification and symbol decoding performance has been achieved in a jointly optimal manner. Various extensions of the results in this study can be considered. One interesting extension is to address the uncertainty in the prior probabilities of the modulation schemes by employing the restricted Bayesian framework [26], [39]. In addition,

parameters such as fading coefficient, time/frequency offset can be incorporated into the signal model to investigate a more general scenario in which these parameters are treated as unknown deterministic or random quantities. As another interesting future work, it is possible to study joint modulation classification and symbol detection problem considering MIMO systems.

REFERENCES

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[2] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications.* Hoboken, NJ, USA: Wiley, 2015.

[3] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Blind modulation classification: A concept whose time has come," in *Proc. IEEE Sarnoff Symp.*, Apr. 2005, pp. 223–228.

[4] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 4, pp. 455–469, Jul. 2011.

[5] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, vol. 2, Oct. 2000, pp. 670–674.

[6] M. L. D. Wong and A. K. Nandi, "Semi-blind algorithms for automatic classification of digital modulation schemes," *Digit. Signal Process.*, vol. 18, no. 2, pp. 209–227, 2008.

[7] J. A. Sills, "Maximum-likelihood modulation classification for PSK/QAM," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, vol. 1, 1999, pp. 217–220.

[8] E. Kanterakis and W. Su, "Modulation classification in MIMO systems," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Nov. 2013, pp. 35–39.

[9] B. Dulek, O. Ozdemir, P. K. Varshney, and W. Su, "Distributed maximum likelihood classification of linear modulations over nonidentical flat block-fading Gaussian channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 724–737, Feb. 2015.

[10] F. Hameed, O. A. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.

[11] Q. Shi and Y. Karasawa, "Noncoherent maximum likelihood classification of quadrature amplitude modulation constellations: Simplification, analysis, and extension," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1312–1322, Apr. 2011.

[12] H. Sarieddeen, M. M. Mansour, L. M. A. Jalloul, and A. Chehab, "Efficient near optimal joint modulation classification and detection for MU-MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3706–3710.

[13] A. A. Tadaion, M. Derakhtian, S. Gazor, and M. R. Aref, "Likelihood ratio tests for PSK modulation classification in unknown noise environment," in *Proc. Can. Conf. Electr. Comput. Eng.*, May 2005, pp. 151–154.

[14] B. Dulek, "Online hybrid likelihood based modulation classification using multiple sensors," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4984–5000, Aug. 2017.

[15] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.

[16] J. L. Xu, W. Su, and M. Zhou, "Software-defined radio equipped with rapid modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1659–1667, May 2010.

[17] N. E. Lay and A. Polydoros, "Per-survivor processing for channel acquisition, data detection and modulation classification," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Oct. 1994, pp. 1169–1173.

[18] A. Abdi, O. A. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, vol. 1, Oct./Nov. 2004, pp. 211–217.

[19] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

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

[21] J. Zhang, D. Cabric, F. Wang, and Z. Zhong, "Cooperative modulation classification for multipath fading channels via expectation-maximization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6698–6711, Oct. 2017.

[22] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 434–444, May 1968.

[23] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4215–4229, Jul. 2012.

[24] G. H. Jajamovich, A. Tajer, and X. Wang, "Minimax-optimal hypothesis testing with estimation-dependent costs," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6151–6165, Dec. 2012.

[25] S. Li and X. Wang, "Optimal joint detection and estimation based on decision-dependent Bayesian cost," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2573–2586, May 2016.

[26] B. Dulek, "A restricted Bayes approach to joint detection and estimation under prior uncertainty," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1767–1782, Aug. 2018.

[27] Y. Yılmaz, S. Li, and X. Wang, "Sequential joint detection and estimation: Optimum tests and applications," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5311–5326, Oct. 2016.

[28] N. Weinberger and N. Merhav, "Codeword or noise? Exact random coding exponents for joint detection and decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5077–5094, Sep. 2014.

[29] N. Weinberger and N. Merhav, "Channel detection in coded communication," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6364–6392, Oct. 2017.

[30] S. Bayram, B. Dulek, and S. Gezici, "Joint detection and decoding in the presence of prior information with uncertainty," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1602–1606, Nov. 2016.

[31] A. Gomaa, L. M. A. Jalloul, M. M. Mansour, K. Gomadam, and D. Tujkovic, "Max-log-MAP optimal MU-MIMO receiver for joint data detection and interferer modulation classification," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1389–1392, Jul. 2016.

[32] H. Sarieddeen, M. M. Mansour, L. Jalloul, and A. Chehab, "High order multi-user MIMO subspace detection," *J. Signal Process. Syst.*, vol. 90, no. 3, pp. 305–321, Mar. 2018.

[33] H. Sarieddeen, M. M. Mansour, and A. Chehab, "Modulation classification via subspace detection in MIMO systems," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 64–67, Jan. 2017.

[34] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov-Smirnov test," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.

[35] P. Urriza, E. Rebeiz, and D. Cabric, "Optimal discriminant functions based on sampled distribution distance for modulation classification," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1885–1888, Oct. 2013.

[36] B. Dulek, O. Ozdemir, P. K. Varshney, and W. Su, "Modulation discovery over arbitrary additive noise channels based on the Richardson-Lucy algorithm," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 756–760, Jun. 2014.

[37] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, "Fold-based Kolmogorov–Smirnov modulation classifier," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 1003–1007, Jul. 2016.

[38] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY, USA: Springer, 2005.

[39] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.

Ertan Kazikli received the B.S. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2012, and the M.S. degree in computer science from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2015. He is currently pursuing the Ph.D. degree in electrical and electronics engineering with Bilkent University. His research interests include statistical signal processing, communication theory, and visible light communications.

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 in the industry. From 2012 to 2013, he was a Post-Doctoral Research Associate with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. Since 2014, he has been with the Department of Electrical and Electronics Engineering, Hacettepe University, where he is currently an Associate Professor. His research interests include statistical signal processing, detection and estimation theory, and communication theory. He is an Editor of the *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 was with the Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. Since 2007, he has been with the Department of Electrical and Electronics Engineering, Bilkent University, where he is currently a Professor. His research interests include detection and estimation theory, wireless communications, and localization systems. Among his publications in the above mentioned areas is the book *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols* (Cambridge University Press, 2008). He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE WIRELESS COMMUNICATIONS LETTERS, and the *Journal of Communications and Networks*.