Distributionally Robust Collaborative Beamforming in D2D Relay Networks with Interference Constraints

Shimin Gong, Sissi Xiaoxiao Wu, Anthony Man-Cho So, and Xiaoxia Huang

Abstract – In this paper, we consider a device-to-device (D2D) network underlying a cellular system wherein the densely deployed D2D user devices can act as wireless relays for a distant transceiver pair. We aim to devise a beamforming strategy for the relays that maximizes the data rate of the distant transceiver while satisfying interference constraints at the cellular receivers. Towards that end, we first formulate a beamforming problem whose solution is robust against the channel uncertainties in the relay-destination hop. Motivated by practical observations, we assume that the random channels in this hop follow unimodal distributions and propose a novel unimodal distributionally robust model to capture the channel uncertainties. Then, we extend the formulation so that it can also guard against the channel uncertainty in the source-relay hop under the worst-case robust model. The resulting robust beamforming problem is generally non-convex and intractable. Therefore, we design an iterative algorithm, which is based on solving semidefinite programs, to find an approximate solution to it. Simulation results show that under mild conditions, our robust model significantly improves the throughput of D2D relay transmissions when compared to conventional robust models that merely rely on the channels' moment information. It also outperforms the Bernstein-type inequality-based convex approximation, which assumes that the channel follows a Gaussian distribution.

Index Terms – Distributional uncertainty, unimodality, robust beamforming, D2D relay network

Shimin Gong and Xiaoxia Huang are with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. Email:{sm.gong,xx.huang}@siat.ac.cn; Sissi Xiaoxiao Wu and Anthony Man-Cho So are with the Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong. Email:{xxwu,manchoso}@se.cuhk.edu.hk.

I. INTRODUCTION

Recently, device-to-device (D2D) communication has been viewed as a promising technology to increase data rate, extend network coverage, and reduce power consumption in cellular networks [1]–[3]. It is mainly proposed to offload the data traffic between the cellular base station (CBS) and the cellular user equipment (CUE) within or beyond the cellular coverage. In this paper, instead of emphasizing the performance improvement in the cellular system, we focus on D2D communications and aim to improve the quality-of-service (QoS) of a D2D network within the coverage of a CBS. This study is especially advantageous for social-related applications such as interactive gaming, file sharing, video streaming, etc., in which the access to local content is more frequent than that to the core network via the CBS.

The dense deployment of D2D user equipment (DUE) allows us to achieve spatial diversity by employing multiple DUE as collaborative relays. It has been shown in [4], [5] that the use of relays significantly improves the network performance while causing insignificant increase in the end-to-end delay. However, it may not be optimal for all the relays to transmit at their peak power [6]. Thus, a distributed beamforming strategy that can adapt to the varying channel conditions at different relays is needed in order to achieve optimal relay performance. Moreover, the dense DUE make the spectrum access more crowded. An economical and spectrally efficient solution is to let the DUE share the same spectrum with the cellular system [7]. To avoid mutual interference among the DUE and the CUE while enhancing the throughput of the DUE relays, an efficient power control strategy is of crucial importance. In [8], each DUE's transmit power is separately controlled while the aggregate interference to the CUE is restricted by limiting the number of collaborative relays. In [9], the authors considered an ad hoc D2D network, in which the DUE's opportunistic channel access ensures the interference to the cellular system is kept below a certain threshold. However, most of the existing works rely on the unrealistic assumption that the channel information is precisely known. In practice, due to quantization errors, processing delay, or the lack of communications between the CUE and DUE, the channel estimation is usually unreliable and subject to random errors.

Knowing that channel uncertainty is inevitable, it is imperative to develop a *quantitative model* to characterize the varying channels, so that proactive strategies can be devised to avoid sharp performance deterioration. To this end, a growing effort has been devoted to model the channel uncertainty and formulate the corresponding robust beamforming problems; see, e.g., [10] for a

survey on robust power control and beamforming in cognitive radio networks. Broadly speaking, there are three different types of channel uncertainty models. The *stochastic (STO) model* assumes that the channel estimate is a random variable following a specific distribution. For example, assuming Rician or Rayleigh distributed channels, the robust beamforming problem in [11] imposes average QoS requirements on both the DUE and CUE. In [12], the authors considered energy efficiency maximization in 5G systems wherein the channel estimates are assumed to be Gaussian distributed. In [13], the channel distribution is assumed to be t-distributed instead of Gaussian. To enable a more flexible control, some works formulate the QoS requirements in a probabilistic manner, thus giving rise to chance-constrained robust beamforming problems; see, e.g., [14]–[17]. The *worst-case robust (WCR) model* stipulates that the channel error lies in a compact convex set and requires the robust beamformer to achieve the best performance even for the worst error; see, e.g., [18], [19]. Hence, robust beamforming with the WCR model is usually formulated as a max-min problem, which is then tackled by convex reformulation or approximation. For example, in [18], the authors proposed a joint resource allocation and power control problem for the DUE relays, in which a guard function is used to approximate each perturbed resource constraint. In [19], the authors showed that the max-min beamforming problem can be efficiently solved for a general class of convex uncertainty sets. By leveraging the well-known $\mathcal{S}$-lemma [20], a tighter and efficiently solvable approximation can be derived by transforming the max-min problem into a semidefinite program (SDP). Lastly, the *distributionally robust (DRO) model* assumes that partial information about the random channel estimate, such as moment statistics, is available. The DRO model was proposed in [21] and recently employed to study robust beamforming by leveraging the first- and second-order moments of the random channel [22]. However, such an approach is still over-conservative, as the worst-case channel distribution in the DRO model is shown to be discrete and is hardly observed in practice [23].

The above discussion motivates us to impose additional requirements on the shape or structure of the channel distribution. Empirically, we find that channel distributions are generally smooth, symmetric, unimodal, or even similar to some known patterns; see, e.g., [13], [24]. Such *structural information* can be extracted and exploited to mitigate the conservatism of the DRO model. In this work, we focus on unimodality and propose, for the first time in the context of wireless communications, the *unimodal distributionally robust (UDR) model* to improve the relays' beamforming design. We consider the scenario where there are uncertainties in both the source-relay and relay-destination channels in a two-hop half-duplex relay transmission scheme,

and we aim to maximize the DUE's SNR subject to the relays' power budget constraint and the CUE's interference constraints. Under this setting, the uncertain channels are coupled with each other, thereby presenting a great challenge in the analysis of the stochastic properties of the DUE's SNR and CUE's interference. It is worth noting that the aforementioned scenario has been studied under the WCR model (i.e., no structural information about the channel estimate is used) in a very recent work [25]. In particular, the authors showed that the corresponding robust beamforming problem admits a biquadratic formulation and can be tackled by semidefinite relaxation techniques [26]. In this work, we depart from the approach in [25] and tackle the robust beamforming problem under a mixed UDR-WCR model by processing the channel uncertainties in both hops separately. The main contributions of this paper are as follows:

1) *Unimodal distributionally robust (UDR) model*: Motivated by the observations that many channel distributions are unimodal, we introduce the UDR model, in which the random channel is assumed to follow a unimodal distribution with known first- and second-order moments. To the best of our knowledge, our work is the first to adopt unimodality in the modeling of channel uncertainty. By leveraging a generalized notion of unimodality [27], we parameterize a class of unimodal distributions by a scalar $\alpha > 0$. As we shall see, a larger value of α implies more conservative channel modeling. In particular, the UDR model degenerates to the conventional DRO model when $\alpha \to \infty$.
2) *Beamforming algorithm for the UDR model:* As a first step towards accounting for the channel uncertainties in both source-relay and relay-destination hops and to avoid a highly complex biquadratic formulation similar to that in [25], we consider the setting where the source-relay channel is known and the relay-destination channels are characterized by the UDR model, and we formulate a robust beamforming problem whose objective is to maximize the DUE's data rate while satisfying the relays' power budget and CUE's UDR (with respect to the uncertainties in the relay-destination channels) interference constraints. Since the UDR interference constraints are non-convex in general, we derive safe approximations of them and propose a tractable iterative procedure that tackles the safe approximation problem by solving a sequence of DRO beamforming problems.
3) *Robustness against uncertain channels in both hops:* To ensure robustness against the uncertain channels in both source-relay and relay-destination hops, we assume that the uncertainty in the source-relay channel is characterized by the WCR model and formulate

a mixed UDR-WCR beamforming problem by extending the formulation developed in item 2 above. We then design a tractable algorithm that uses the iterative procedure developed in item 2 above as building block to tackle the resulting formulation. Simulation results show that our algorithm significantly improves the DUE's data rate over the DRO model. It also outperforms the approach in [17], which is based on the strong assumption that the channel follows a Gaussian distribution and uses a Bernstein-type inequality and semidefinite relaxation techniques to tackle the resulting chance-constrained robust beamforming problem.

A preliminary version of this work has appeared as a conference paper [28]. The rest of this paper is organized as follows. In Section II, we introduce the system model and the UDR model to describe channel uncertainties. In Section III, we first present our two-step approach to formulating the problem of finding a beamforming design that maximizes the DUE's SNR while satisfying the relays' power budget and the CUE's mixed UDR-WCR interference constraints. Then, we develop iterative procedures to find an approximate solution to the aforementioned robust relay beamforming problem. Lastly, we present performance evaluations and conclusions in Sections IV and V, respectively.

II. Problem Setup and Preliminaries

Consider a D2D network with densely deployed user devices underlying a downlink cellular system, in which each DUE transmitter (DTx), DUE relay, and DUE receiver (DRx) are equipped with a single antenna. As shown in Fig. 1, the direct link from DTx to DRx is not available due to the long distance between them or limited power at the DTx. The data transmission is assisted by a set of DUE relays, denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$. The DUE relays employ the amplify-and-forward (AF) scheme, which is known for its short processing delay and simplicity in system deployment [4]. As the D2D network and the DUE relays share the same spectrum with the cellular system, the relays' transmissions may introduce interference to the CUE, which we denote by $\mathcal{K} = \{1, 2, \ldots, K\}$. The data reception at the CUE will not be interrupted if the interference from the DUE relays is less than a pre-defined threshold. The downlink cellular transmissions from the CBS may also introduce interference to the D2D network. We assume that the power level of the noise and the CBS' interference at each DUE is known and fixed during the D2D transmissions. In particular, each individual DUE can estimate the power level in a silence period before initiating data transmissions. This assumption is also valid when D2D

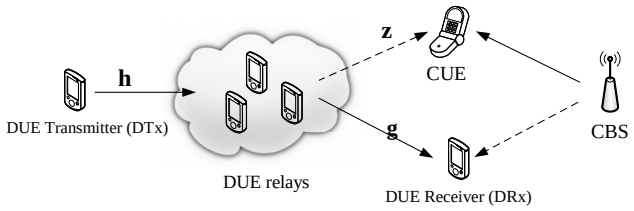

Fig. 1: System model.

communications occur in the cell-edge region, where the interference from the CBS is small and can be viewed as background noise with a fixed power level.

A. Signal Model

The data transmission follows a two-hop relay scheme. In the first hop, the complex channel from the DTx to the relays is denoted by $\mathbf{h} \triangleq [h_1, h_2, \ldots, h_N]^T \in \mathbb{C}^N$; in the second hop, the complex channels from the relays to the DRx and from the relays to CUE $k \in \mathcal{K}$ are denoted by $\mathbf{g} \triangleq [g_1, g_2, \ldots, g_N]^T \in \mathbb{C}^N$ and $\mathbf{z}_k \triangleq [z_{1k}, z_{2k}, \ldots, z_{Nk}]^T \in \mathbb{C}^N$, respectively. In the first hop, the DTx broadcasts a symbol $s \in \mathbb{C}$ to its nearby relays. Assuming that all channels are frequency flat and block fading [29], the received signal at the relays is given by $\mathbf{m} = \mathbf{h}s + \boldsymbol{\sigma}$, where $\boldsymbol{\sigma} \triangleq [\sigma_1, \sigma_2, \ldots, \sigma_N]^T$ denotes the complex Gaussian noise with zero mean and unit variance (i.e., $\boldsymbol{\sigma} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ with $\mathbf{I}$ being the identity matrix). For notational simplicity, we ignore the time index in the signal model.

In the second hop, the relays forward the signal $\mathbf{m}$ weighted by a complex amplifying coefficient $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T \in \mathbb{C}^N$. Thus, the received signal at the DRx is

$$y = \mathbf{g}^H(\mathbf{w} \circ \mathbf{h})s + \mathbf{g}^H(\mathbf{w} \circ \boldsymbol{\sigma}) + v_d, \tag{1}$$

where $\circ$ denotes the Hadamard product and $v_d \sim \mathcal{CN}(0, 1)$ is the noise at the DRx. Here, $\mathbf{w}$ is viewed as the relays' beamforming vector. The first term in (1) contains the desired signal, while the second term is the noise forwarded by the relays. Obviously, it is not optimal for all the relays to transmit at their peak power, as doing so will also amplify and forward the noise. Assuming $\mathbb{E}[|s|^2] = 1$, the SNR at the DRx is given by

$$\gamma(\mathbf{w}) = \frac{\left|\mathbf{g}^H(\mathbf{w} \circ \mathbf{h})\right|^2}{1 + |\mathbf{g} \circ \mathbf{w}|^2} = \frac{\mathbf{w}^H \mathbf{A} \mathbf{w}}{1 + \mathbf{w}^H \mathbf{B} \mathbf{w}},$$

where $\mathbf{A} = (\mathbf{g} \circ \mathbf{h})(\mathbf{g} \circ \mathbf{h})^H$ and $\mathbf{B} = \mathrm{D}(\mathbf{g}^* \circ \mathbf{g})$ denotes a diagonal matrix with the diagonal elements given by the vector $\mathbf{g}^* \circ \mathbf{g}$. Here, $\mathbf{g}^*$ represents the complex conjugate of $\mathbf{g}$.

To avoid excessive interference to the CUE, the relays' beamforming has to restrict the aggregate interference to the CUE. Since the transmit power at relay-n is given by $|w_n|^2(1+|h_n|^2)$, the aggregate interference at the CUE $k \in \mathcal{K}$ can be expressed as

$$\phi_k(\mathbf{w}) = \mathbf{z}_k^H \mathrm{D}(\mathbf{c})\mathrm{D}(\mathbf{w}^* \circ \mathbf{w})\mathbf{z}_k = \mathbf{w}^H \mathrm{D}(\mathbf{c})\mathrm{D}(\mathbf{z}_k^* \circ \mathbf{z}_k)\mathbf{w},$$

where $\mathbf{c} = \mathbf{1} + \mathbf{h}^* \circ \mathbf{h}$. Our goal is to maximize the DUE's SNR $\gamma(\mathbf{w})$ subject to the CUE's interference constraints and the relays' power budget constraint; i.e.,

$$\max_{\mathbf{w} \in P} \{\gamma(\mathbf{w}) : \phi_k(\mathbf{w}) \le \bar{\phi}_k, \ \forall k \in \mathcal{K}\}, \tag{2}$$

where $\bar{\phi}_k$ is the interference threshold of the CUE $k \in \mathcal{K}$ and represents the CUE's sensitivity or tolerance to interference, and $P \triangleq \{\mathbf{w} \in \mathbb{C}^N \mid \mathbf{c}^T(\mathbf{w}^* \circ \mathbf{w}) \le \bar{p}\}$ is the set of feasible beamforming vectors with $\bar{p}$ being the relays' maximum transmit power. We can also add a power budget constraint for each individual relay, but this will not give rise to new design challenges in our problem formulation. For ease of presentation, we only consider one sum power constraint in the feasible set P.

B. Distributional Uncertainty

The optimal solution to (2) relies on exact channel information; i.e., the channel-dependent matrix coefficients $\mathbf{A}$ and $\mathbf{B}$ must be acquired in advance. In the two-hop relay scheme, the source-relay channel $\mathbf{h}$ can be estimated by the relays in real time. The error due to quantization can be confined in a small set and thus the conventional WCR model performs well. However, to estimate the channels in the second hop, the feedback information from the receivers is not always available due to the lack of communications with the CUE. Without regular information exchange with the receivers, the channel estimate could be far different from its actual value. In this case, the WCR model will lead to severe performance degradation. This motivates us to look for a less conservative uncertainty model for the channels in the second hop.

The uncertainty model for the channels $\mathbf{g}$ and $\mathbf{z}$ in the second hop is built upon the well-known DRO model [22]. Specifically, we assume that the random channel $\mathbf{x} \in \{\mathbf{g}, \mathbf{z}\}$ follows a certain distribution $\mathbb{P}_\mathbf{x}$ whose probability density function (pdf) is not precisely known. As the first- and second-order moments of the random channel can be estimated rather easily, we stipulate that

$$\mathbb{P}_\mathbf{x} \in \mathcal{P}(\mathbf{u}_\mathbf{x}, \mathbf{S}_\mathbf{x}) \subset \mathcal{P}_\infty, \tag{3}$$

where $\mathcal{P}_\infty$ denotes the set of all probability distributions and $\mathcal{P}(\mathbf{u_x}, \mathbf{S_x})$ is the set of distributions with first- and second-order moments given by $\mathbf{u_x}$ and $\mathbf{S_x}$, respectively. Unfortunately, analytical results in [23] show that the DRO model (3) is still too pessimistic, as the set $\mathcal{P}(\mathbf{u_x}, \mathbf{S_x})$ includes discrete or multi-modal distributions, which rarely arise in practice.

In this work, we mitigate the conservatism of the DRO model by imposing additional requirement on the structure or shape of the channel distributions in $\mathcal{P}(\mathbf{u_x}, \mathbf{S_x})$. Such structural information can be easily extracted from channel measurements. In particular, we require that the distribution $\mathbb{P}_\mathbf{x}$ to be unimodal, which is commonly observed in wireless communications and signal processing. For example, the recent work [13] shows that the interference power can be viewed as a unimodal random variable. The measurements in [30] reveal that the shadow fading signal converges to a unimodal distribution that resembles a Gaussian. In [31], the authors studied the convexity properties of symbol error rate under a general class of unimodal noise power densities. Intuitively, if a distribution is unimodal, then its pdf has a single local maximum (referred to as the mode). In particular, the pdf is non-increasing along the rays emanating from the mode. Some well-known distributions, such as Gaussian, Rayleigh, and Rician distributions, are unimodal. It should be noted that these distributions differ significantly in terms of the structure or shape of their pdfs.

C. Characterizing Unimodality

By incorporating unimodality into the DRO model, we can remove those distributions that are not commonly encountered in wireless communications and signal processing from the set (3), thereby leading to a more practical distributional uncertainty model for the channels in the second hop. To exploit this model in algorithm design, however, we first need an analytical characterization of unimodality for *multivariate* distributions. This can be achieved using the notion of α-unimodality [27].

Definition 1 (**α-Unimodality [27]**): *Suppose that $\mathbb{P} \in \mathcal{P}_\infty$ has a continuous pdf f on $\mathbb{R}^N$. We say that $\mathbb{P}$ is α-unimodal with mode $\mathbf{0}$ for some $\alpha > 0$ if $t^{N-\alpha} f(t\mathbf{x})$ is non-increasing in $t > 0$ for $\mathbf{x} \neq \mathbf{0}$. The set of all α-unimodal distributions with mode $\mathbf{0}$ is denoted as $\mathcal{P}_\alpha$.*

When $\alpha = N$, the dimension of the random variable $\mathbf{x}$, α-unimodality implies that f is non-increasing along the rays emanating from its mode, which coincides with our intuition for univariate unimodal distributions. When $\alpha > N$, the pdf f may increase along the rays emanating

from its mode, but the weighted pdf $t^{N-\alpha}f(t\mathbf{x})$ can be kept non-increasing in $t > 0$. Through different choices of α, we can control the rate of increase of f and hence α-unimodality can be viewed as a generalized characterization of unimodal distributions. By choosing $\alpha = N$, the set $\mathcal{P}_\alpha$ will contain most of the practically observed distributions [27], such as the Gaussian and Rayleigh distributions. Another practical way of choosing α is to use an online learning process. Specifically, each relay samples its channels for some time period and fits an empirical pdf f_e using the sampled data. Then, we can choose α as the *minimum* value such that $t^{N-\alpha}f_e(t\mathbf{x})$ is non-increasing in $t > 0$ for $\mathbf{x} \neq \mathbf{0}$. For a different set of sampled data, the empirical pdf f_e could be different and so is the parameter α. The relay can thus repeat the sampling procedure for several times and choose α to be the *largest* one.

For the complex channel $\mathbf{x} = \text{Re}(\mathbf{x}) + i\text{Im}(\mathbf{x})$, we assume that its real and imaginary parts (denoted by $\text{Re}(\mathbf{x})$ and $\text{Im}(\mathbf{x})$, respectively) are independent. Thus, the distribution of $\mathbf{x}$ can be characterized by the joint distribution $f(\mathbf{x}) = f_R(\text{Re}(\mathbf{x}))f_I(\text{Im}(\mathbf{x}))$, where f, f_R, and f_I denote the pdfs of $\mathbf{x}$, $\text{Re}(\mathbf{x})$, and $\text{Im}(\mathbf{x})$, respectively. This relationship can be leveraged to establish the unimodality of a complex channel's distribution.

Proposition 1: *Suppose that the distributions of Re*$(\mathbf{x})$ *and Im*$(\mathbf{x})$ *are* $(\alpha/2)$*-unimodal for some* $\alpha > 0$*. Then, the distribution of* $\mathbf{x} =$ *Re*$(\mathbf{x}) + i$*Im*$(\mathbf{x}) \in \mathbb{C}^N$ *is* α*-unimodal; i.e.,* $\mathbb{P}_\mathbf{x} \in \mathcal{P}_\alpha$.

The proof of Proposition 1 easily follows by verifying that $t^{2N-\alpha}f_\mathbf{x}(t\mathbf{x})$ is non-increasing in $t > 0$. Note that we have $t^{2N-\alpha}f_\mathbf{x}(t\mathbf{x}) = t^{N-\alpha/2}f_R(t\text{Re}(\mathbf{x}))t^{N-\alpha/2}f_I(t\text{Im}(\mathbf{x}))$. Since $\mathbb{P}_{\text{Re}(\mathbf{x})}$, $\mathbb{P}_{\text{Im}(\mathbf{x})} \in \mathcal{P}_{\alpha/2}$, both $t^{N-\alpha/2}f_R(t\text{Re}(\mathbf{x}))$ and $t^{N-\alpha/2}f_I(t\text{Im}(\mathbf{x}))$ are non-increasing in $t > 0$, which implies that $t^{2N-\alpha}f_\mathbf{x}(t\mathbf{x})$ is also non-increasing in $t > 0$. Hence, we have $\mathbb{P}_\mathbf{x} \in \mathcal{P}_\alpha$, as desired. We now define the *unimodal distributionally robust (UDR) model* for the complex channel $\mathbf{x}$ as

$$\mathbb{P}_\mathbf{x} \in \mathcal{P}_\mathbf{x}^\alpha(\mathbf{u}_\mathbf{x}, \mathbf{S}_\mathbf{x}) \triangleq \mathcal{P}_\mathbf{x}(\mathbf{u}_\mathbf{x}, \mathbf{S}_\mathbf{x}) \cap \mathcal{P}_\alpha. \tag{4}$$

By tuning the parameter α, the parameterized set $\mathcal{P}_\mathbf{x}^\alpha(\mathbf{u}_\mathbf{x}, \mathbf{S}_\mathbf{x})$ (or $\mathcal{P}_\mathbf{x}^\alpha$ for short) provides the flexibility to model the uncertainty of $\mathbf{x}$ under different channel conditions.

III. Robust Relay Beamforming with Interference Constraints: Formulations and Algorithms

A. Robustness against Uncertainties in the Relay-Destination Channels under the UDR Model

With the preparations in the previous section, we are now ready to study the robust counterparts of the beamforming problem (2). We begin by considering the case where the source-relay

channel $\mathbf{h}$ is known but the relay-destination channels $\mathbf{g}, \mathbf{z}$ are stochastic and drawn from the distributions $\mathbb{P}_{\mathbf{g}}$ and $\mathbb{P}_{\mathbf{z}}$, respectively. Since both the SNR $\gamma(\mathbf{w})$ and the interference $\phi_k(\mathbf{w})$ become stochastic, we formulate the following UDR counterpart of (2):

$$\max_{\mathbf{w}\in P} \quad \min_{\mathbb{P}_{\mathbf{g}}\in\mathcal{P}^{\alpha}_{\mathbf{g}}} \ \frac{\mathbb{E}_{\mathbb{P}_{\mathbf{g}}}[\mathbf{w}^H\mathbf{A}\mathbf{w}]}{1+\mathbb{E}_{\mathbb{P}_{\mathbf{g}}}[\mathbf{w}^H\mathbf{B}\mathbf{w}]} \tag{5a}$$

$$\text{s.t.} \quad \max_{\mathbb{P}_{\mathbf{z}_k}\in\mathcal{P}^{\alpha}_{\mathbf{z}_k}} \ \mathbb{P}_{\mathbf{z}_k}\left(\mathbf{z}_k^H\mathbf{\Lambda}_{\mathbf{w}}\mathbf{z}_k \geq \bar{\phi}_k\right) \leq \eta, \quad \forall k \in \mathcal{K}. \tag{5b}$$

Here, we define $\mathbf{\Lambda}_{\mathbf{w}} \triangleq \mathrm{D}(\mathbf{c})\mathrm{D}(\mathbf{w}^*\circ\mathbf{w})$. By designing the relays' beamformer $\mathbf{w}$, we maximize the DUE's worst-case SNR while limiting the CUE's *worst-case interference violation probability* (with respect to all distributions with the given structural and moment information) below the probability target η. The UDR beamforming problem (5) is difficult to solve optimally and its tractability depends on our ability to come up with a convex reformulation/approximation of the non-convex chance constraint (5b). This involves finding a tractable upper bound on the probability that a unimodally distributed random vector lies outside a given *ellipsoid*. It is worth noting that the work [27] gives an upper bound on the probability that a unimodally distributed random vector lies outside a *polyhedron* and shows that such an upper bound admits an exact SDP formulation. However, it is clear that such a result is insufficient for our purposes.

To simplify the formulation (5), we first introduce the SNR target $\rho \geq 0$ as an auxiliary variable and observe that (5a) is equivalent to maximizing ρ subject to the following QoS constraint at the DUE transceiver pair:

$$\rho + \max_{\mathbb{P}_{\mathbf{g}}\in\mathcal{P}^{\alpha}_{\mathbf{g}}} \mathbb{E}_{\mathbb{P}_{\mathbf{g}}}\left[\mathbf{g}^H\left(\rho\mathrm{D}(\mathbf{w}^*\circ\mathbf{w}) - \mathbf{q}\mathbf{q}^H\right)\mathbf{g}\right] \leq 0, \tag{6}$$

where we denote $\mathbf{q} = \mathrm{D}(\mathbf{h})\mathbf{w}$ for simplicity. Note that (6) involves only the second-order moment $\mathbf{S}_g$ and not the particular structure of $\mathbb{P}_{\mathbf{g}}$. Hence, we can simplify (6) as

$$\rho + \rho\mathrm{Tr}(\mathrm{D}(\mathbf{w}^*\circ\mathbf{w})\mathbf{S}_{\mathbf{g}}) - \mathbf{q}^H\mathbf{S}_{\mathbf{g}}\mathbf{q} \leq 0. \tag{7}$$

Note that for a fixed ρ, the constraint (7) defines a linear inequality with respect to the matrix variable $\mathbf{w}\mathbf{w}^H$, which can be easily handled by the semidefinite relaxation technique. This will be detailed shortly.

Next, let us consider the probabilistic interference constraint (5b). Define

$$e_k(\mathbf{z}) = \mathbf{1}(\mathbf{z}^H\mathbf{\Lambda}_{\mathbf{w}}\mathbf{z} \geq \bar{\phi}_k), \tag{8}$$

where $\mathbf{1}(\cdot)$ is the indicator function. In other words, $e_k(\mathbf{z})$ indicates whether the interference threshold of the CUE $k \in \mathcal{K}$ is violated. Given the relays' beamformer $\mathbf{w}$, we further define

$$B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \;= \max_{\mathbb{P}_{\mathbf{z}_k} \in \mathcal{P}_{\mathbf{z}_k}^{\alpha}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}_k}}[e_k(\mathbf{z}_k)] \tag{9}$$

to be the worst-case interference violation probability on the LHS of (5b). Since the constraint $B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \leq \eta$ has no known closed-form convex equivalence, we shall focus on deriving a tractable upper bound on $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ in the sequel. To proceed, we first study a reformulation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ in the special case where $\alpha = \infty$. After getting some insights from this special case, we then derive a safe approximation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ for the general case where α is finite.

1) A Special Case of the UDR Model: As α approaches infinity, the UDR model (4) degenerates into the DRO model (3). In this case, we can reformulate $B_k^{\infty}(\mathbf{w}, \mathbf{h})$ as follows:

Proposition 2: *The constraint $B_k^{\infty}(\mathbf{w}, \mathbf{h}) \leq \eta$ is equivalent to*

$$0 \;\geq \min_{\mathbf{M}_k,\, \nu_k} \quad \left(-\nu_k + \frac{1}{\eta}\mathbf{Tr}(\mathbf{\Sigma}_{\mathbf{z}_k}\mathbf{M}_k)\right) \tag{10a}$$

$$\text{s.t.} \quad \mathbf{M}_k \succeq \begin{bmatrix} \mathbf{\Lambda}_{\mathbf{w}} & \mathbf{0} \\ \mathbf{0} & \nu_k - \bar{\phi}_k \end{bmatrix}, \tag{10b}$$

$$\mathbf{M}_k \succeq \mathbf{0}, \quad \nu_k \geq 0, \tag{10c}$$

where $\mathbf{\Sigma}_{\mathbf{z}_k} = \begin{bmatrix} \mathbf{S}_{\mathbf{z}_k} & \mathbf{u}_{\mathbf{z}_k} \\ \mathbf{u}_{\mathbf{z}_k}^T & 1 \end{bmatrix}$ *denotes the second-order moment matrix of channel* $\mathbf{z}_k$.

The proof of Proposition 2 follows a rather standard Lagrangian approach and the details can be found in [22]. For a fixed $\mathbf{w}$, problem (10) is an SDP and thus provides a convex equivalence for the worst-case interference violation probability (9). Upon replacing (5b) by (10), we see that the optimal beamformer $\mathbf{w}^{\star}$ can be obtained from the following optimization problem:

$$\max_{\mathbf{w} \in P,\, \rho,\, \mathbf{M}_k,\, \nu_k} \quad \rho \tag{11a}$$

$$\text{s.t.} \quad \rho + \rho \mathrm{Tr}(\mathbf{D}(\mathbf{w}^* \circ \mathbf{w})\mathbf{S}_{\mathbf{g}}) - \mathbf{q}^H \mathbf{S}_{\mathbf{g}} \mathbf{q} \leq 0, \tag{11b}$$

$$\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{z}_k}\mathbf{M}_k) \leq \nu_k \eta, \tag{11c}$$

$$\mathbf{M}_k \succeq \begin{bmatrix} \mathbf{D}(\mathbf{c})\mathbf{D}(\mathbf{w}^* \circ \mathbf{w}) & \mathbf{0} \\ \mathbf{0} & \nu_k - \bar{\phi}_k \end{bmatrix}, \tag{11d}$$

$$\mathbf{M}_k \succeq \mathbf{0},\, \nu_k \geq 0, \quad \forall k \in \mathcal{K}. \tag{11e}$$

Unfortunately, problem (11) is still non-convex because the constraints (11b) and (11d) are quadratic in $\mathbf{w}$. To circumvent this difficulty, we first introduce the rank-one matrix $\mathbf{W} = \mathbf{w}\mathbf{w}^H$ and note that $\mathbf{q}^H\mathbf{S_g}\mathbf{q} = \mathbf{h}^H(\mathbf{W} \circ \mathbf{S_g})\mathbf{h}$ and $\mathrm{D}(\mathbf{w}^* \circ \mathbf{w}) = \mathbf{\Delta}(\mathbf{W})$, where $\mathbf{\Delta}(\mathbf{W})$ is the diagonal matrix obtained by setting all off-diagonal elements of $\mathbf{W}$ to zero. Then, by dropping the non-convex rank-one constraint on $\mathbf{W}$, we obtain the following semidefinite relaxation of the UDR beamforming problem (11):

$$
\begin{aligned}
(\mathrm{SUB}): \max_{\mathbf{W},\,\rho,\,\mathbf{M}_k,\,\nu_k} \quad & \rho && \text{(12a)}\\
\text{s.t.} \quad & \bar{\mathbf{S}}_{\mathbf{g}}(\rho, \mathbf{W}) - \mathbf{h}^H(\mathbf{W} \circ \mathbf{S_g})\mathbf{h} \le 0, && \text{(12b)}\\
& \mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{z}_k}\mathbf{M}_k) \le \nu_k\eta, && \text{(12c)}\\
& \mathbf{M}_k \succeq \begin{bmatrix} \mathrm{D}(\mathbf{c})\mathbf{\Delta}(\mathbf{W}) & \mathbf{0} \\ \mathbf{0} & \nu_k - \bar{\phi}_k \end{bmatrix}, && \text{(12d)}\\
& \mathrm{Tr}(\mathrm{D}(\mathbf{c})\mathbf{W}) \le \bar{p}, && \text{(12e)}\\
& \mathbf{W} \succeq \mathbf{0},\ \mathbf{M}_k \succeq \mathbf{0},\ \nu_k \ge 0, \quad \forall k \in \mathcal{K}, && \text{(12f)}
\end{aligned}
$$

where $\bar{\mathbf{S}}_{\mathbf{g}}(\rho, \mathbf{W}) \triangleq \rho + \rho\mathrm{Tr}(\mathbf{\Delta}(\mathbf{W})\mathbf{S_g})$.

Problem (SUB) can be solved optimally using a bisection approach. Specifically, for a given target SNR ρ, checking the feasibility of (12b)–(12f) is an SDP, which can be solved efficiently by interior-point algorithms [32]. Hence, we can increase or decrease ρ depending on whether the constraints (12b)–(12f) are satisfied or not, until convergence is achieved. The computational complexity of (SUB) depends on the complexity of checking the feasibility of (12b)–(12f) for a fixed ρ and the number of bisection steps. A detailed analysis will be provided in Section III-B. Here, we note that if the optimal solution $\mathbf{W}^\star$ to (SUB) happens to be of rank one, then the optimal rank-one beamformer $\mathbf{w}^\star$ to (11) can be extracted from $\mathbf{W}^\star$ by eigen-decomposition. Otherwise, we can extract an approximate rank-one solution to (11) from $\mathbf{W}^\star$ by the Gaussian randomization method [26]. Extensive experiments show that $\mathbf{W}^\star$ is of rank one most of the time. A similar observation has also been made in [22].

2) The General UDR Model: Now, let us consider the general UDR model. By definition of α-unimodality, a smaller α implies more stringent structural requirement; i.e., $\mathcal{P}_{\alpha_1} \supset \mathcal{P}_{\alpha_2}$ for any $\alpha_1 \ge \alpha_2 > 0$. Thus, by tuning the value of α, we can construct a family of robust formulations that capture the uncertainty under different channel conditions. To obtain a concrete representation of $\mathcal{P}_\alpha$ for finite values of α, we need the following notion:

Definition 2 (Radial α-Unimodality [27]): *For any $\alpha > 0$ and $\mathbf{x} \in \mathbb{R}^N$, the radial α-unimodal distribution supported on the line segment $[\mathbf{0}, \mathbf{x}] \subset \mathbb{R}^N$, denoted by $\delta^{\alpha}_{[\mathbf{0},\mathbf{x}]}(\cdot)$, is an α-unimodal distribution with the property that $\delta^{\alpha}_{[\mathbf{0},\mathbf{x}]}([\mathbf{0}, t\mathbf{x}]) = t^{\alpha}$ for $t \in [0, 1]$.*

Proposition 3: *Given $\mathbb{P} \in \mathcal{P}_{\alpha}$, there exists a unique $\mathbb{m} \in \mathcal{P}_{\infty}$ such that $\mathbb{P}(\cdot) = \int_{\mathbb{R}^N} \delta^{\alpha}_{[\mathbf{0},\mathbf{x}]}(\cdot)\mathbb{m}(d\mathbf{x})$.*

The proof of Proposition 3 for real distributions can be found in [27]. It can be extended to complex distributions by applying Proposition 1 if we assume that the real and imaginary parts of the random vector follow independent $(\alpha/2)$-unimodal distributions.

Proposition 3 asserts that any α-unimodal distribution can be uniquely represented by a mixture of radial α-unimodal distributions. Thus, it provides a mapping between an α-unimodal distribution $\mathbb{P} \in \mathcal{P}_{\alpha}$ and its mixture distribution $\mathbb{m} \in \mathcal{P}_{\infty}$. Since there is no unimodality constraint on $\mathbb{m}$, we can reformulate (9) as the following conventional moment problem:

Proposition 4: *For $0 \leq \alpha < +\infty$, we have*

$$B^{\alpha}_{k}(\mathbf{w}, \mathbf{h}) = \max_{\mathbb{m} \in \mathcal{P}_{\infty}\left(\frac{1+\alpha}{\alpha}\mathbf{u}_{\mathbf{z}_k}, \frac{2+\alpha}{\alpha}\mathbf{S}_{\mathbf{z}_k}\right)} \mathbb{E}_{\mathbb{m}}\left[p^{\alpha}_{k}(\mathbf{x})\right], \tag{13}$$

where $p^{\alpha}_{k}(\mathbf{x}) \triangleq \int_{\mathbb{R}^n} e_k(\mathbf{z})\delta^{\alpha}_{[\mathbf{0},\mathbf{x}]}(d\mathbf{z})$ and $e_k(\mathbf{z})$ is defined in (8)*.*

The proof of Proposition 4 follows by applying Proposition 3 to problem (9). The details are omitted here for conciseness. Proposition 4 implies that once we find some $\mathbb{m}$ that is feasible for (13), we can construct an α-unimodal distribution $\mathbb{P}_{\mathbf{z}_k}$ that is feasible for (9). Unfortunately, the conventional Lagrangian approach does not yield a tractable dual of (13). Thus, we resort to deriving a tractable upper bound on (13), which can then be used to construct a safe approximation of (5b).

Proposition 5: *The following problem*[1] *yields an upper bound on* $B_k^{\alpha}(\mathbf{w},\mathbf{h})$*:*

$$\max_{\mathbf{Y},\mathbf{y},\lambda,\tau\geq 0} \quad \lambda-\tau \tag{14a}$$

$$\text{s.t.} \quad \begin{bmatrix} \bar{\mathbf{S}}_{\mathbf{z}_k} & \bar{\mathbf{u}}_{\mathbf{z}_k} \\ \bar{\mathbf{u}}_{\mathbf{z}_k}^T & 1 \end{bmatrix} - \begin{bmatrix} \mathbf{Y} & \mathbf{y} \\ \mathbf{y}^T & \lambda \end{bmatrix} \succeq \mathbf{0}, \tag{14b}$$

$$\begin{bmatrix} \mathbf{Y} & \mathbf{y} \\ \mathbf{y}^T & \lambda \end{bmatrix} \succeq \mathbf{0}, \tag{14c}$$

$$\tau^2\big(\mathbf{Tr}(\mathbf{\Lambda_w Y})\big)^{\alpha} \geq \lambda^{\alpha+2}\bar{\phi}_k^{\alpha}. \tag{14d}$$

The proof of Proposition 5 is relegated to Appendix A. The main difficulty in solving problem (14) lies in the non-convex constraint (14d). To tackle this constraint, let us assume that $\alpha>0$ is an integer and introduce an auxiliary variable $s\geq 0$ such that

$$\big(\mathrm{Tr}(\mathbf{\Lambda_w Y})\big)^{\alpha} s^{\alpha} \;\geq\; \lambda^{2\alpha}\bar{\phi}_k^{\alpha}, \tag{15a}$$

$$\tau^2\lambda^{\alpha-2}s^{2^l-\alpha} \;\geq\; s^{2^l}, \tag{15b}$$

where $l=\lceil \log_2(\alpha)\rceil$ is the smallest integer satisfying $2^l-\alpha\geq 0$. By imposing mild conditions on α, we can derive an equivalent convex formulation of (15a)–(15b) as follows:

Proposition 6: *Let* $\alpha\geq 2$ *be an integer. By introducing the auxiliary variables* $t_{2^i+m}\geq 0$*, where* $i=0,1,\ldots,l-1$*, the constraints (15a)–(15b) are equivalent to*

$$\begin{bmatrix} \mathbf{Tr}(\mathbf{\Lambda_w Y}) & \lambda\bar{\phi}_k^{\frac{1}{2}} \\ \lambda\bar{\phi}_k^{\frac{1}{2}} & s \end{bmatrix} \;\succeq\; \mathbf{0}, \tag{16a}$$

$$\begin{bmatrix} t_{2^{i+1}+2m} & t_{2^i+m} \\ t_{2^i+m} & t_{2^{i+1}+2m+1} \end{bmatrix} \;\succeq\; \mathbf{0}, \tag{16b}$$

$$t_1 \;\geq\; s \tag{16c}$$

for $m=0,1,\ldots,2^i-1$ *and* $i=0,1,\ldots,l-1$*, where*

$$t_{2^l+m}=\begin{cases} \tau & \textit{for } m=0,1, \\ \lambda & \textit{for } m=2,\ldots,\alpha-1, \\ s & \textit{for } m=\alpha,\ldots,2^l-1. \end{cases} \tag{17}$$

[1]With a slight abuse of notation, we drop the subscript k on $(\mathbf{Y},\mathbf{y},\lambda,\tau)$ in (14) for notational simplicity.

The proof of Proposition 6 is relegated to Appendix B. Now, note that given a fixed $\mathbf{w}$, the constraints (16a)–(16b) are linear matrix inequalities. Thus, problem (14) can be solved efficiently by leveraging the SDP reformulation (16). In particular, we obtain the following bi-level optimization formulation of the general UDR beamforming problem:

$$(\mathrm{GUB}): \max_{\rho,\, \mathbf{w}\in P} \quad \rho \tag{18a}$$

$$\text{s.t.} \quad \rho + \rho \mathrm{Tr}(\mathbf{D}(\mathbf{w}^* \circ \mathbf{w})\mathbf{S_g}) - \mathbf{q}^H \mathbf{S_g} \mathbf{q} \leq 0, \tag{18b}$$

$$\eta \geq \max\{\lambda_k - \tau_k \,:\, (14b), (14c), \text{ and } (16)\}, \quad \forall k \in \mathcal{K}. \tag{18c}$$

On the upper-level of problem (GUB), we optimize the rank-one beamformer $\mathbf{w}$ to maximize the target SNR ρ. On the lower-level of problem (GUB), we compute the worst-case interference violation probability η of all the CUE. The constraint (18b) ensures the SNR performance at the destination DUE receiver. However, due to the coupling between $\mathbf{w}$ and $\mathbf{Y}$ in the non-convex matrix inequality (16a), the upper- and lower-level problems are not separable, thus rendering problem (GUB) intractable. To circumvent this difficulty, we make the following observation, which reveals a connection between (GUB) and (SUB):

Proposition 7: *The upper bound on $B_k^\alpha(\mathbf{w}, \mathbf{h})$ derived by (14) is increasing in $\alpha > 0$.*

We relegate the proof in Appendix C. Proposition 7 implies that (SUB) produces the largest evaluation of $B_k^\alpha(\mathbf{w}, \mathbf{h})$ and thus can be viewed as a restriction of (GUB); i.e., any $\mathbf{w}$ that is feasible for (SUB) is also feasible for (GUB). Moreover, since the added structural information in (GUB) mitigates the conservatism in channel estimation, (GUB) will achieve a better SNR performance than that of (SUB) for the same beamformer $\mathbf{w}$. This observation leads to the novel idea of approximating (GUB) by iteratively checking the feasibility of (SUB). Specifically, the algorithm starts from a beamformer $\mathbf{w}$ that is feasible for (SUB). Then, we fix $\mathbf{w}$ and solve the lower-level problem (14). Once we obtain an upper bound on $B_k^\alpha(\mathbf{w}, \mathbf{h})$, we check the feasibility of (18c) and then update $\mathbf{w}$ accordingly. The whole procedure is presented in Algorithm 1. Given the CUE's probability threshold η (e.g., $\eta = 0.1$), we initialize $(\rho, \mathbf{W})$ in line 3 of Algorithm 1 by solving (SUB) with the initial threshold $\tilde{\eta} = \eta$. The rank-one solution $\mathbf{w}$ can be extracted from $\mathbf{W}$ by eigen-decomposition or the Gaussian randomization method. Then, fixing the beamformer $\mathbf{w}$, we check the feasibility of (18c) in problem (GUB). If $B_k^\alpha(\mathbf{w}, \mathbf{h})$ is smaller (resp. larger) than the target probability limit η, we increase (resp. decrease) the value of $\tilde{\eta}$ by bisection and then

Algorithm 1 UDR Relay Beamforming against Uncertainties in the Relay-Destination Hop

initialize $\tilde{\eta}_{\min} = \eta$, $\tilde{\eta}_{\max} = 1$, and $\tilde{\eta} = \eta$
while $|\tilde{\eta}_{\max} - \tilde{\eta}_{\min}| \geq \varepsilon$
  find $(\rho, \mathbf{W})$ by solving (SUB) with the target $\tilde{\eta}$
  extract the rank-one beamformer $\mathbf{w}$ from $\mathbf{W}$
  evaluate $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ by solving the SDP (14)
  if $B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \leq \eta$ for all $k \in \mathcal{K}$
    update $\tilde{\eta}_{\min} \leftarrow \tilde{\eta}$ **else** $\tilde{\eta}_{\max} \leftarrow \tilde{\eta}$
  end if
  update $\tilde{\eta} \leftarrow (\tilde{\eta}_{\max} + \tilde{\eta}_{\min})/2$
end while
return $(\rho, \mathbf{w})$

solve (SUB) again to update $(\rho, \mathbf{W})$. At convergence, Algorithm 1 always returns a rank-one solution to (GUB), which provides a lower bound on the optimal value of problem (5) for any integer $\alpha \geq 2$.

B. Robustness against Uncertainty in the Source-Relay Channel under the WCR Model

We now extend the UDR formulation (5) to incorporate uncertainty in the source-relay channel $\mathbf{h}$. Observe that we can write $\mathbf{h} = \bar{\mathbf{h}} + \Delta_{\mathbf{h}}$, where $\bar{\mathbf{h}}$ is the nominal channel estimate and $\Delta_{\mathbf{h}}$ is the estimation error. Since the DUE relays can perform frequent channel estimation, we typically have $\|\Delta_{\mathbf{h}}\| \ll \|\bar{\mathbf{h}}\|$. This motivates us to adopt the following conventional WCR model for $\mathbf{h}$:

$$\mathbb{U} \triangleq \{\mathbf{h} \in \mathbb{C}^N \mid \mathbf{h} = \bar{\mathbf{h}} + \Delta_{\mathbf{h}} \text{ and } \|\Delta_{\mathbf{h}}\| \leq \delta\}, \tag{19}$$

where $\delta > 0$ is a given parameter. Using the uncertainty set $\mathbb{U}$, we can formulate the following analog of (GUB), which ensures robustness against the uncertainties in both the source-relay channel $\mathbf{h}$ and the relay-destination channels $\mathbf{g}$, $\mathbf{z}$:

$$(\mathrm{GUB})' : \max_{\rho, \mathbf{w} \in P} \quad \rho \tag{20a}$$

$$\text{s.t.} \quad \bar{\mathbf{S}}_{\mathbf{g}}(\rho, \mathbf{w}) - \mathbf{q}^H \mathbf{S}_{\mathbf{g}} \mathbf{q} \leq 0, \quad \forall\, \mathbf{h} \in \mathbb{U}, \tag{20b}$$

$$\eta \geq \max_{\mathbf{h} \in \mathbb{U}} B_k^{\alpha}(\mathbf{w}, \mathbf{h}), \quad \forall k \in \mathcal{K}. \tag{20c}$$

Here, we define $\bar{\mathbf{S}}_{\mathbf{g}}(\rho, \mathbf{w}) \triangleq \rho + \rho \mathrm{Tr}(\mathrm{D}(\mathbf{w}^* \circ \mathbf{w})\mathbf{S}_{\mathbf{g}})$. It is straightforward to show that Proposition 7 still holds for $\max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\mathbf{w}, \mathbf{h})$, which implies that we can solve $(\mathrm{GUB})'$ in the same way as (GUB). We summarize the procedure in Algorithm 2. Our proposed procedure relies on the tractability of the safe approximation in Step 1 and the feasibility check in Step 2. Note that Algorithm 2 performs bisection over $\tilde{\eta}$ to search for the feasible beamformer and thus converges linearly. In the sequel, we show that both Steps 1 and 2 of Algorithm 2 admit an efficient implementation.

Algorithm 2 Robust Relay Beamforming against Uncertainties in Both Hops

Step 1: Solve a safe approximation of $(\mathrm{GUB})'$ by replacing (20c) with

$$\tilde{\eta} \geq \max_{\mathbf{h}\in\mathbb{U}} B_k^{\infty}(\mathbf{w}, \mathbf{h}), \quad k \in \mathcal{K}.$$

Step 2: Let $\widetilde{\mathbf{w}}$ be the beamformer obtained in Step 1. Check if the following constraints are satisfied:

$$\eta \geq \max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\widetilde{\mathbf{w}}, \mathbf{h}), \quad k \in \mathcal{K}.$$

Step 3: If not, update $\tilde{\eta}$ using the bisection method and go to Step 1 until convergence is achieved.

1) Solving the Safe Approximation in Step 1: By applying the techniques developed in Section III, we obtain the following representation of the safe approximation in Step 1:

$$\max_{\mathbf{w},\,\rho,\,\mathbf{M}_k,\,\nu_k} \quad \rho \tag{21a}$$

$$\text{s.t.} \quad \mathbf{c}^T(\mathbf{w}^* \circ \mathbf{w}) \leq \bar{p}, \quad \forall\, \mathbf{h} \in \mathbb{U}, \tag{21b}$$

$$\bar{\mathbf{S}}_{\mathbf{g}}(\rho, \mathbf{w}) - \mathbf{q}^H \mathbf{S}_{\mathbf{g}} \mathbf{q} \leq 0, \quad \forall\, \mathbf{h} \in \mathbb{U}, \tag{21c}$$

$$\mathbf{M}_k \succeq \begin{bmatrix} \mathrm{D}(\mathbf{c})\mathrm{D}_{\mathbf{w}} & \mathbf{0} \\ \mathbf{0} & \nu_k - \bar{\phi}_k \end{bmatrix}, \quad \forall\, \mathbf{h} \in \mathbb{U}, \tag{21d}$$

$$\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{z}_k} \mathbf{M}_k) \leq \nu_k \eta, \tag{21e}$$

$$\mathbf{M}_k \succeq \mathbf{0}, \ \nu_k \geq 0, \quad \forall\, k \in \mathcal{K}. \tag{21f}$$

Here, we define $\mathrm{D}_{\mathbf{w}} = \mathrm{D}(\mathbf{w}^* \circ \mathbf{w})$ for simplicity. The power budget constraint (21b) and the SNR requirement (21c) are quadratic in estimation error $\Delta_{\mathbf{h}}$. As such, they can be transformed

into linear matrix inequalities using the $\mathcal{S}$-lemma. Indeed, after some manipulations, the power budget constraint (21b) can be written as

$$(\bar{\mathbf{h}}+\Delta_{\mathbf{h}})^H \mathrm{D}_{\mathbf{w}}(\bar{\mathbf{h}}+\Delta_{\mathbf{h}})+\mathbf{w}^H\mathbf{w}\le \bar{p}, \quad \forall \|\Delta_{\mathbf{h}}\|^2\le \delta^2.$$

By the $\mathcal{S}$-lemma, there exists a $t_1\ge 0$ such that

$$\left[\begin{array}{cc} t_1\mathbf{I}-\mathrm{D}_{\mathbf{w}} & -\mathrm{D}_{\mathbf{w}}\bar{\mathbf{h}} \\ -\bar{\mathbf{h}}^H\mathrm{D}_{\mathbf{w}} & \bar{p}-\mathbf{w}^H\mathbf{w}-\bar{\mathbf{h}}^H\mathrm{D}_{\mathbf{w}}\bar{\mathbf{h}}-t_1\delta^2 \end{array}\right]\succeq \mathbf{0}. \tag{22}$$

Upon setting $\bar{\mathbf{c}}\triangleq \mathbf{1}+\bar{\mathbf{h}}^*\circ\bar{\mathbf{h}}$ and $\mathbf{W}=\mathbf{w}\mathbf{w}^H$, we obtain the following linear matrix inequality representation:

$$\left[\begin{array}{cc} t_1\mathbf{I}-\boldsymbol{\Delta}(\mathbf{W}) & -\boldsymbol{\Delta}(\mathbf{W})\bar{\mathbf{h}} \\ -\bar{\mathbf{h}}^H\boldsymbol{\Delta}(\mathbf{W}) & \bar{p}-\mathrm{Tr}(\mathrm{D}(\bar{\mathbf{c}})\mathbf{W})-t_1\delta^2 \end{array}\right]\succeq \mathbf{0}. \tag{23}$$

In a similar fashion, we have the following linear matrix inequality representation of (21c):

$$\left[\begin{array}{cc} t_2\mathbf{I}+\mathbf{T}_{\mathbf{w}} & \mathbf{T}_{\mathbf{w}}\bar{\mathbf{h}} \\ \bar{\mathbf{h}}^H\mathbf{T}_{\mathbf{w}} & \bar{\mathbf{h}}^H\mathbf{T}_{\mathbf{w}}\bar{\mathbf{h}}-\bar{\mathbf{S}}_{\mathbf{g}}(\rho,\mathbf{W})-t_2\delta^2 \end{array}\right]\succeq \mathbf{0}, \tag{24}$$

where $t_2\ge 0$ and $\mathbf{T}_{\mathbf{w}}\triangleq \mathbf{W}\circ\mathbf{S}_{\mathbf{g}}$.

Next, we analyze the more challenging semi-infinite constraint (21d). Let

$$\bar{\mathbf{M}}_k\triangleq \mathbf{M}_k-\mathrm{D}\left(\left[\begin{array}{c}\mathbf{w}^*\circ\mathbf{w}\\ \nu_k-\bar{\phi}_k\end{array}\right]\right).$$

Then, we can express (21d) as $\bar{\mathbf{M}}_k\succeq \left[\begin{array}{cc}\mathrm{D}(\mathbf{h}^*\circ\mathbf{h}) & \mathbf{0}\end{array}\right]^H\left[\begin{array}{cc}\mathrm{D}_{\mathbf{w}} & \mathbf{0}\end{array}\right]$. Let $\mathbf{P}=\left[\begin{array}{cc}\mathrm{D}_{\mathbf{w}} & \mathbf{0}\end{array}\right]$ and $\mathbf{K}_k=\bar{\mathbf{M}}_k-\left[\begin{array}{cc}\mathrm{D}(\bar{\mathbf{h}}^*\circ\bar{\mathbf{h}}) & \mathbf{0}\end{array}\right]^H\mathbf{P}$. We require that

$$\mathbf{K}_k\succeq \left[\begin{array}{c}\mathrm{D}(\Delta_{\mathbf{h}}^*\circ\Delta_{\mathbf{h}}+2\bar{\mathbf{h}}^*\circ\Delta_{\mathbf{h}})\\ \mathbf{0}\end{array}\right]\mathbf{P} \tag{25}$$

holds for any $\Delta_{\mathbf{h}}$ satisfying $\|\Delta_{\mathbf{h}}\|\le\delta$. Define $\mathbf{X}=\left[\begin{array}{cc}\mathrm{D}(\Delta_{\mathbf{h}}) & \mathbf{0}\end{array}\right]^H$. Clearly, we have $\|\mathbf{X}\|\triangleq\sqrt{\lambda_{\max}(\mathbf{X}\mathbf{X}^H)}\le\delta$. Hence, a *sufficient condition* to ensure (25) is given by

$$\mathbf{K}_k\succeq 2\left[\begin{array}{cc}\mathrm{D}(\bar{\mathbf{h}}+\frac{\delta}{2}\mathbf{1}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\mathbf{X}\mathbf{P}, \quad \forall\|\mathbf{X}\|\le\delta. \tag{26}$$

To process (26), we need the following result:

Lemma 1 (Petersen's Lemma [33]): *Given matrices* $\mathbf{P}$, $\mathbf{Q}$, *and* $\mathbf{K}$ *with* $\mathbf{K}=\mathbf{K}^H$, *we have*

$$\mathbf{K}\succeq \mathbf{Q}^H\mathbf{X}^H\mathbf{P}+\mathbf{P}^H\mathbf{X}\mathbf{Q}, \quad \forall\|\mathbf{X}\|\le\delta \tag{27}$$

if and only if there exists a $t \geq 0$ *such that*

$$\begin{bmatrix} \mathbf{K} - t\mathbf{Q}^H\mathbf{Q} & -\delta\mathbf{P}^H \\ -\delta\mathbf{P} & t\mathbf{I} \end{bmatrix} \succeq \mathbf{0},$$

or equivalently $\mathbf{K} - t\mathbf{Q}^H\mathbf{Q} - (\delta^2/t)\mathbf{P}^H\mathbf{P} \succeq \mathbf{0}$.

By Petersen's Lemma, the constraint (26) is equivalent to

$$\begin{bmatrix} \mathbf{K}_k - \begin{bmatrix} t_{3,k}\mathrm{D}_\delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & -\delta \begin{bmatrix} \mathrm{D}_{\mathbf{w}} \\ \mathbf{0} \end{bmatrix} \\ -\delta \begin{bmatrix} \mathrm{D}_{\mathbf{w}} & \mathbf{0} \end{bmatrix} & t_{3,k}\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \tag{28}$$

where $t_{3,k} \geq 0$ and $\mathrm{D}_\delta \triangleq \mathrm{D}(\bar{\mathbf{h}}^* + \frac{\delta}{2}\mathbf{1})\mathrm{D}(\bar{\mathbf{h}} + \frac{\delta}{2}\mathbf{1})$. Upon setting $\mathbf{W} = \mathbf{w}\mathbf{w}^H$, we obtain the following representation of (28):

$$\mathbf{M}_k' \triangleq \begin{bmatrix} \mathbf{M}_k & \mathbf{0}_{(n+1)\times n} \\ \mathbf{0}_{n\times(n+1)} & t_{3,k}\mathbf{I}_n \end{bmatrix} \succeq \begin{bmatrix} \mathrm{D}(\bar{\mathbf{c}})\boldsymbol{\Delta}(\mathbf{W}) + t_{3,k}\mathrm{D}_\delta & \mathbf{0}_{n\times 1} & \delta\boldsymbol{\Delta}(\mathbf{W}) \\ \mathbf{0}_{1\times n} & \nu_k - \bar{\phi}_k & \mathbf{0}_{1\times n} \\ \delta\boldsymbol{\Delta}(\mathbf{W}) & \mathbf{0}_{n\times 1} & \mathbf{0}_{n\times n} \end{bmatrix}. \tag{29}$$

Since (28) is a safe approximation of (21d), we conclude that the optimal value of problem (21) is lower bounded by the objective value associated with an optimal rank-one solution to the following problem:

$$\begin{aligned} (\mathrm{SUB})' : \max_{\substack{\mathbf{W},\,\rho,\,\mathbf{M}_k,\,\nu_k \\ t_1,\,t_2,\,t_{3,k}}} \quad & \rho \\ \text{s.t.} \quad & \text{(21e), (23), (24), and (29)}, \\ & t_1 \geq 0,\ t_2 \geq 0,\ \mathbf{W} \succeq \mathbf{0}, \\ & t_{3,k} \geq 0,\ \mathbf{M}_k \succeq \mathbf{0},\ \nu_k \geq 0, \quad \forall k \in \mathcal{K}. \end{aligned}$$

Similar to problem (SUB), we can solve problem $(\mathrm{SUB})'$ using the bisection method. Indeed, for a fixed ρ, checking the feasibility of $(\mathrm{SUB})'$ is an SDP. At the convergence of the bisection method, we can extract from an optimal $\mathbf{W}^\star$ a feasible beamformer using either eigen-decomposition or the Gaussian randomization.

2) Feasibility Check in Step 2: Step 2 of Algorithm 2 requires the evaluation of $\max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\mathbf{w},\mathbf{h})$ for some integer $\alpha \geq 2$ and comparing the result to the probability target η. By Propositions 5 and 6, an upper bound on $\max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\mathbf{w},\mathbf{h})$ is given by

$$\begin{aligned} \max_{\mathbf{Y},\mathbf{y},\lambda,\tau\geq 0} \quad & \lambda-\tau \\ \text{s.t.} \quad & \text{(14b)–(14c), (16b)–(16c), and (17)}, \\ & \max_{\mathbf{h}\in\mathbb{U}} \ \mathrm{Tr}(\mathbf{\Lambda}_{\mathbf{w}}\mathbf{Y}) \geq \frac{\lambda^2}{s}\bar{\phi}_k. \end{aligned} \tag{30}$$

Observe that for a given $\mathbf{Y}$, the maximization of $\mathrm{Tr}(\mathbf{\Lambda}_{\mathbf{w}}\mathbf{Y})$ with respect to $\mathbf{h}\in\mathbb{U}$ takes the form

$$\max_{\|\Delta_{\mathbf{h}}\|^2\leq\delta^2} (\bar{\mathbf{h}}+\Delta_{\mathbf{h}})^H \mathbf{Y}_{\mathbf{w}} (\bar{\mathbf{h}}+\Delta_{\mathbf{h}}), \tag{31}$$

where $\mathbf{Y}_{\mathbf{w}} \triangleq (\mathrm{D}_{\mathbf{w}}\circ\mathbf{Y}) \succeq \mathbf{0}$. By [34, Theorem 3.2], an optimal solution to (31) can be obtained by solving an SDP.

To solve (30), we employ the alternating optimization (AO) technique. Specifically, we first solve (30) with fixed $\mathbf{h}^{(t)}\in\mathbb{U}$ to obtain $(\mathbf{Y}^{(t)},\mathbf{y}^{(t)},\lambda^{(t)},\tau^{(t)})$. This is equivalent to solving an instance of problem (14). Then, we fix $(\mathbf{Y}^{(t)},\mathbf{y}^{(t)},\lambda^{(t)},\tau^{(t)})$ and solve (31) to get $\mathbf{h}^{(t+1)} = \bar{\mathbf{h}} + \Delta_{\mathbf{h}}^{(t+1)}$. The entire procedure is summarized in Algorithm 3. It is easy to verify that the sequence of objective values $\{\lambda^{(t)}-\tau^{(t)}\}$ is increasing, which allows us to establish the convergence of the AO method using the results in [35].

Algorithm 3 AO Method for Evaluating $\max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\mathbf{w},\mathbf{h})$

1: initialize $\Delta_{\mathbf{h}}^{(t-1)} = \mathbf{0}$ and $\Delta_{\mathbf{h}}^{(t)} = \delta\mathbf{e}$ for $t=1$, where $\mathbf{e}$ is a unit vector
2: **while** $\|\Delta_{\mathbf{h}}^{(t)} - \Delta_{\mathbf{h}}^{(t-1)}\| \geq \epsilon$
3: solve the SDP (14) with fixed $\mathbf{h}^{(t)}$
4: update $\Delta_{\mathbf{h}}^{(t+1)}$ by solving problem (31)
5: $\mathbf{h}^{(t+1)} \leftarrow \bar{\mathbf{h}} + \Delta_{\mathbf{h}}^{(t+1)}$
6: $t \leftarrow t+1$
7: **end while**

So far we showed that Algorithm 2, which involves solving a sequence of SDPs, provides a tractable way of solving a safe approximation of $(\mathrm{GUB})'$. Let us now give a rough analysis of its computational complexity. The complexity of Step 1 depends on the number of iterations required to solve $(\mathrm{SUB})'$ times the complexity of solving an SDP. Given the error

tolerance ϵ in the bisection method, the total number of iterations is on the order of $\log\left(\frac{1}{\epsilon}\right)$. The complexity of solving an SDP with $\tilde{n}$ decision variables and $\tilde{K}$ linear matrix inequalities is $O\left(\left(\sum_{i=1}^{\tilde{K}}\tilde{m}_i\right)^{1/2}\left(\tilde{n}^2\sum_{i=1}^{\tilde{K}}\tilde{m}_i^2+\tilde{n}\sum_{i=1}^{\tilde{K}}\tilde{m}_i^3\right)\right)$ [17], where $\tilde{m}_i$ denotes the dimension of the i-th linear matrix inequality. By simple manipulations, we see that the overall complexity of Step 1 is on the order of $\log\left(\frac{1}{\epsilon}\right)K^{3.5}N^{6.5}$, where K and N denote the number of CUE and the DUE relays, respectively. In Step 2, we need to check the feasibility of the constraint $\eta \geq \max_{\mathbf{h}\in\mathbb{U}} B_k^{\alpha}(\mathbf{w},\mathbf{h})$, which is carried out by Algorithm 3. In each iteration of Algorithm 3, we need to solve the SDP (14) for a fixed $\mathbf{h}^{(t)}$ and problem (31) for a fixed $(\mathbf{Y}^{(t)},\mathbf{y}^{(t)},\lambda^{(t)},\tau^{(t)})$. Simulation results reveal that Algorithm 3 requires few iterations to converge, and the complexity of Step 2 is mainly dominated by that of solving the SDP (14), which is $O\big(KN^{6.5}\big)$. Since the number of bisection steps over $\tilde{\eta}$ used to find a feasible beamformer is on the order of $\log\left(\frac{1}{\epsilon}\right)$, we conclude that the overall complexity of Algorithm 2 is on the order of $\log\left(\frac{1}{\epsilon}\right)K^{3.5}N^{6.5}$.

IV. Simulation Results

In this section, we demonstrate the efficacy of our proposed UDR-based beamforming design and compare it with those obtained from some existing robust models. For simplicity, we consider 3 DUE relays collaboratively amplifying and forwarding the received signals to the DUE receiver. Since the presence of multiple CUE (i.e., $K>1$) will only result in multiple constraints of the form (18c) in problem (GUB) or (20c) in problem $(\mathrm{GUB})'$, we simply consider the case where $K=1$ in our simulations. We assume that the noise at each relay and at the DUE has zero mean and unit variance. Moreover, the channel $\mathbf{h}$ in the first hop is characterized by the WCR model; i.e., $\mathbf{h}$ is assumed to lie in the spherical set (19). On the other hand, the channels $\mathbf{g}$ and $\mathbf{z}$ have α-unimodal pdfs and known first- and second-order moments, but their exact pdfs are unknown.

A. Worst-Case Interference Violation Probability

To better understand the impact of the UDR model on the interference violation probability, we consider a fixed channel $\mathbf{h}$ and examine how the worst-case interference violation probability $B_k^{\alpha}(\mathbf{w},\mathbf{h})$ changes with the parameter α. For the general UDR model, an upper bound on $B_k^{\alpha}(\mathbf{w},\mathbf{h})$ can be obtained by solving problem (14) with a fixed beamformer $\mathbf{w}$, assuming $\alpha\geq 2$ is an integer (see Proposition 6). Without loss of generality, we set $\mathbf{w}$ to be the solution to (SUB). In Fig. 2, we show the upper bound on $B_k^{\alpha}(\mathbf{w},\mathbf{h})$ with different values of α and the

CUE's interference requirements $(\bar{\phi}, \eta)$. In Fig. 2(a), we observe that the upper bound decreases as the CUE interference threshold $\bar{\phi}$ increases. Despite the randomness in the channel estimates, the aggregate interference $\phi = \mathbf{z}^T \mathbf{\Lambda}_{\mathbf{w}} \mathbf{z}$ received by the CUE tends to concentrate around its mean. Hence, when $\bar{\phi}$ becomes larger, the probability that the deviation from the mean exceeds $\bar{\phi}$ becomes smaller. In Fig. 2(b), we observe that the upper bound on $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ is increasing in α, which is consistent with the conclusion of Proposition 7. In particular, a larger α implies less stringent structural requirement, thus resulting in a more conservative (i.e., larger) evaluation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$.

B. Price of Robustness for Channel Uncertainties

Of course, the UDR model can also be used to characterize the uncertainty of $\mathbf{h}$. However, we are currently unable to derive a tractable approximation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ when all the channels $\mathbf{h}$, $\mathbf{g}$, and $\mathbf{z}$ are unimodal random variables. We circumvent this difficulty by ignoring the structural information on the distribution of $\mathbf{h}$ and simply characterizing its uncertainty by the conventional WCR model (19). Then, we can derive an upper bound on $\max_{\mathbf{h} \in \mathbb{U}} B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ and ensure the solution returned by Algorithm 2 is robust against the uncertain channels in both hops. It is worth noting that such an approach can still perform well when the DUE relays estimate the channel frequently. In Fig. 3(a), we show the upper bound on $\max_{\mathbf{h} \in \mathbb{U}} B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ by varying the normalized error bound $\delta_n = \delta / \|\mathbf{h}\|$, where the relays' beamformer $\mathbf{w}_i$ is obtained by solving problem $(\mathrm{SUB})'$. The results confirm our intuition that the uncertainty in the channel

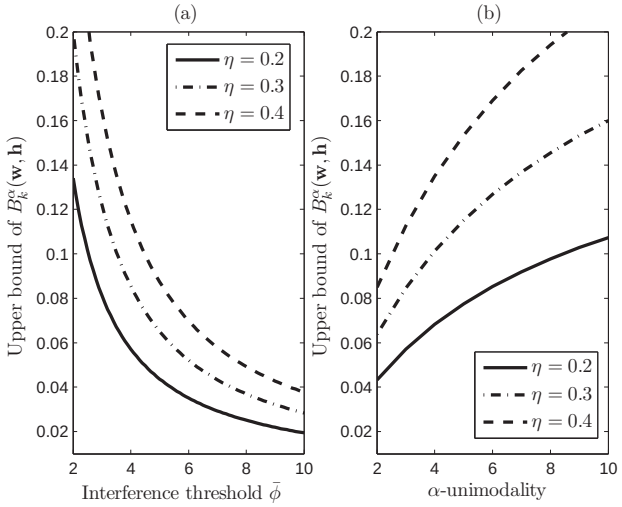

Fig. 2: Upper bounds on the interference violation probability $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$.

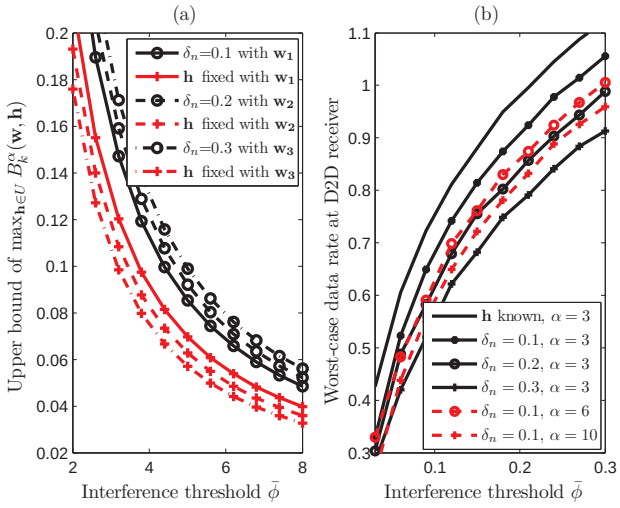

Fig. 3: The UDR model is less conservative than the DRO model.

$\mathbf{h}$ inevitably leads to a more conservative (i.e., larger) estimation of the interference violation probability. Consequently, the degradation in the data rate in Fig. 3(b) can be viewed as the price of robustness paid to guard against the uncertainty in $\mathbf{h}$, which is an increasing function of the error bound δ_n; i.e., the performance gap becomes larger when δ_n increases. We also observe in Fig. 3(b) that the data rate decreases as α increases, which verifies that the upper bound on $\max_{\mathbf{h}\in\mathbb{U}} B_k^\alpha(\mathbf{w},\mathbf{h})$ derived in (30) is an increasing function of α (see Proposition 7).

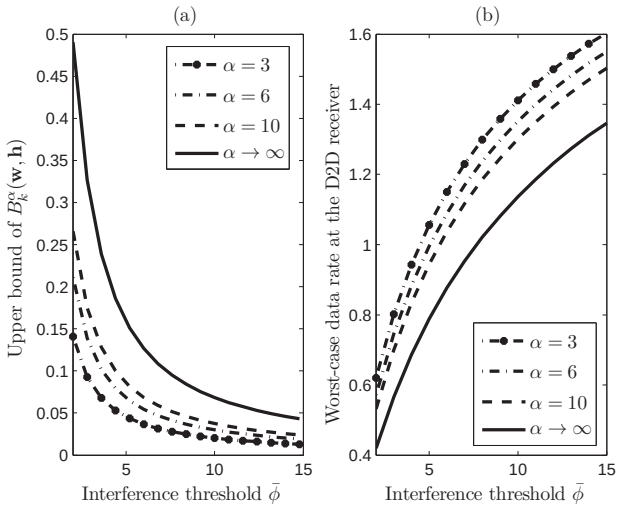

Fig. 4: The UDR model is less conservative than the DRO model.

TABLE I: Simulation results with different $\bar{\phi}$

$\bar{\phi}$	5	6	7	8	9	10
DRO	0.71%	0.65%	0.67%	0.65%	0.60%	0.66%
BER	2.01%	1.98%	2.09%	2.05%	2.00%	2.08%
UDR	6.43%	6.32%	6.46%	6.34%	6.31%	6.29%

C. Relationship between the UDR and DRO Models

The UDR model provides a more practical channel uncertainty model by leveraging the notion of α-unimodality. To verify the performance of the UDR model, we compare it with some existing models. In the sequel, we fix the source-relay channel $\mathbf{h}$ and focus on the uncertain channels $\mathbf{g}$ and $\mathbf{z}$ in the relay-destination hop, which are characterized by the proposed UDR model. As $\alpha \to \infty$, the UDR model degenerates into the DRO model [22], and we can evaluate the exact value of $B_k^{\infty}(\mathbf{w}, \mathbf{h})$ by the SDP (10). For the general UDR model, we only have a safe approximation of (5b). A comparison between the UDR and DRO models is presented in Fig. 4, where the DRO model is denoted by $\alpha \to \infty$. We observe that a larger α in the UDR model implies more conservative evaluation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$. This leads to over-protection for the CUE and thus performance loss at the DUE. Due to the lack of structural information, the DRO model significantly overestimates the interference violation probability as shown in Fig. 4(a). The evaluation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h})$ is used to check whether a given beamformer $\mathbf{w}$ is feasible for the interference constraint; i.e., $B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \leq \eta$. The result of the feasibility check then motivates the update of $\mathbf{w}$ in Algorithm 1. In our simulations, we set the probability limit to be $\eta = 0.2$ and the resulting data rate is shown in Fig. 4(b). Though Algorithm 1 is based on heuristics and a safe approximation of the constraint $B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \leq \eta$, it provides the DUE a much higher data rate than that can be achieved by the DRO model.

D. Performance Comparison with the BER Model [17]

The Bernstein-type inequality (BER) model assumes that the channels are Gaussian distributed and relies on a Bernstein-type inequality to develop a safe approximation of $B_k^{\alpha}(\mathbf{w}, \mathbf{h}) \leq \eta$. Note that the Gaussian distribution is α-unimodal for any α greater than the dimension of the channel vector [27]. To compare the UDR and BER models, we first optimize the beamformer $\mathbf{w}$ in both models for the CUE probability limit $\eta = 0.2$ and then check their throughput performance using random realizations of the channels $\mathbf{g}$ and $\mathbf{z}$. Assuming that the uncertain channels $\mathbf{g}$

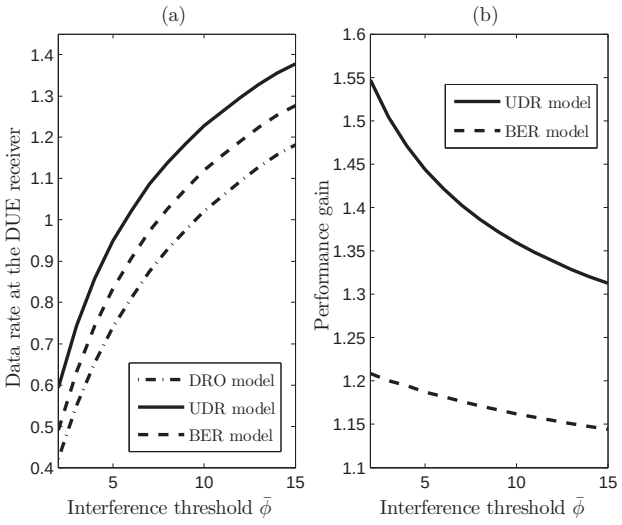

Fig. 5: Throughput performance of three robust models.

and $\mathbf{z}$ are Gaussian distributed, we generate 10^6 channel realizations according to their moment information in each simulation run. For each interference threshold $\bar{\phi}$, we record the CUE's interference violation probabilities in Table I and show the DUE's data rate in Fig. 5. We observe in Table I that the interference violation probabilities do not vary too much with the increase of $\bar{\phi}$, and they do not exceed the prescribed probability limit η. We also notice that the observed interference violation probability in the DRO model is much smaller than that of the BER and UDR models. Again, this is due to the lack of structural information in the DRO model, which leads to the most conservative DUE performance; see also Fig. 5.

A perhaps counter-intuitive observation from Table I is that the UDR model has the highest interference violation probability and consequently a better throughput performance than that of the BER model as shown in Fig. 5. This can be explained in part by the fact that although the BER model demands exact distributional information, the approximation in the Bernstein-type inequality results in an over-estimation of the interference violation probability. As a result, the robust beamformer design in the BER model is more conservative than that of the UDR model. In Fig. 5(b), the performance gain of the UDR (resp. BER) model is defined as the expected ratio $\mathbb{E}[\rho_U/\rho_D]$ (resp. $\mathbb{E}[\rho_B/\rho_D]$), where ρ_D, ρ_U, and ρ_B denote the optimal SNR targets of the DRO, UDR, and BER models, respectively. We observe that the UDR model improves the DUE's performance significantly by imposing additional structural information on the channel distribution.

TABLE II: Simulation results with different α values

α	3	4	5	6	8	10
DRO	0.96%	0.94%	0.97%	1.06%	0.97%	1.13%
BER	1.90%	1.92%	1.91%	1.99%	1.86%	2.02%
UDR	7.56%	6.49%	5.98%	5.37%	4.30%	3.96%
$\mathbb{E}[\rho_B/\rho_D]$	1.151	1.151	1.152	1.153	1.152	1.152
$\mathbb{E}[\rho_U/\rho_D]$	1.556	1.487	1.452	1.415	1.342	1.305

E. Flexibility Achieved by Tuning the Parameter α

By tuning the value of α, the UDR model has the flexibility to model the channel uncertainty under different channel conditions. To show this, we vary the value of α and record some performance metrics of various robust models in Table II. In the simulations, we set $\bar{\phi} = 6$ and $\eta = 0.2$. Note that the DRO and BER models are oblivious to the unimodal structure in the distribution. Thus, we record nearly constant interference violation probabilities in the first two rows of Table II. By the same reason, the performance gain $\mathbb{E}[\rho_B/\rho_D]$ is also a constant for different α. However, when we reduce α and thus impose more stringent structural requirement on the distribution, the UDR model achieves higher SNR by pushing the interference violation probability closer to its target η. Hence, we observe an increasing value of $\mathbb{E}[\rho_U/\rho_D]$ as α decreases. When $\alpha = \infty$, the UDR model degenerates into the DRO model and the performance gain $\mathbb{E}[\rho_U/\rho_D]$ tends to unity. When $\alpha = N$, we achieve a significant performance improvement (around 50%) over the DRO model, as shown in Table II.

V. Conclusion

In this paper, we proposed a two-step approach to the problem of distributionally robust beamforming in a D2D relay network with cellular users. In the first step, we sought for a beamforming design for the relays that is robust against the channel uncertainties in the relay-destination hop and maximizes the DUE's SNR while satisfying the relays' power budget and CUE's interference constraints. Then, in the second step, we extend our formulation in the first step to further guard against the channel uncertainty in the source-relay hop. For the channel uncertainties in the relay-destination hop, we introduced, for the first time in the context of wireless communications, the UDR model to characterize them. Such a model is motivated by the observation that channel distributions are typically unimodal, and this structural information can be obtained without extra effort in the DUE relays' channel estimation. Furthermore, the UDR

model can mitigate the conservatism of other conventional channel uncertainty models such as the WCR and DRO models. As for the channel uncertainty in the source-relay hop, we characterize it using the conventional WCR model, which is reasonable if the DUE relays can perform frequent channel estimation. The aforementioned two-step approach circumvents the complicated coupling of the uncertain channels and facilitates the design of an iterative procedure that can efficiently approximate the optimal beamformer for the DUE relays by solving a sequence of SDPs. Compared with the well-known DRO and BER models, our UDR-based beamforming design is shown to achieve a much higher throughput (typically around 50% performance gain in our simulations). We expect that the UDR model and the solution techniques developed in this paper will find many other applications in robust designs for wireless communications.

APPENDIX

A. *Proof of Proposition 5*

We need to show that for any feasible solution $\mathbb{m}$ to (13), we can construct a feasible solution $(\mathbf{Y}, \mathbf{y}, \lambda, \tau)$ to (14) that achieves at least the same probability bound as $\mathbb{m}$. Let $\mathbb{P}_{\mathbf{z}}$ be a feasible α-unimodal distribution for (9) and $\mathbb{m}$ denote the corresponding mixture distribution in (13). Define $\Xi \triangleq \{\mathbf{x} \,|\, \mathbf{x}^T \mathbf{\Lambda_w} \mathbf{x} < \bar{\phi}_k\}$ and $\bar{\Xi} \triangleq \mathbb{R}^n \setminus \Xi$. Then, we can construct a feasible solution to (14) as follows:

$$\left[\begin{array}{cc} \mathbf{Y} & \mathbf{y} \\ \mathbf{y}^T & \lambda \end{array}\right] = \int_{\bar{\Xi}} \left[\begin{array}{cc} \mathbf{x}\mathbf{x}^T & \mathbf{x} \\ \mathbf{x}^T & 1 \end{array}\right] \mathbb{m}(d\mathbf{x}) \succeq \mathbf{0}, \tag{32a}$$

$$\tau = \int_{\bar{\Xi}} \big(1 - p_k^\alpha(\mathbf{x})\big)\, \mathbb{m}(d\mathbf{x}) \geq 0. \tag{32b}$$

Obviously, we have $\lambda \leq 1$ and

$$\left[\begin{array}{cc} \bar{\mathbf{S}}_{\mathbf{z}_k} & \bar{\mathbf{u}}_{\mathbf{z}_k} \\ \bar{\mathbf{u}}_{\mathbf{z}_k}^T & 1 \end{array}\right] = \int_{\mathbf{x} \in \Xi \cup \bar{\Xi}} \left[\begin{array}{cc} \mathbf{x}\mathbf{x}^T & \mathbf{x} \\ \mathbf{x}^T & 1 \end{array}\right] \mathbb{m}(d\mathbf{x}) \succeq \left[\begin{array}{cc} \mathbf{Y} & \mathbf{y} \\ \mathbf{y}^T & \lambda \end{array}\right].$$

Noting that $p_k^\alpha(\mathbf{x}) = \int_{\mathbb{R}^n} e_k(\mathbf{z}) \delta^\alpha_{[\mathbf{0},\mathbf{x}]}(d\mathbf{z})$, we have

$$1 - p_k^\alpha(\mathbf{x}) = \left\{ \begin{array}{ll} \left(\frac{\bar{\phi}_k}{\mathbf{x}^T \mathbf{\Lambda_w} \mathbf{x}}\right)^{\frac{\alpha}{2}} & \mathbf{x} \in \bar{\Xi} \\ 1, & \mathbf{x} \in \Xi \end{array} \right. ,$$

which implies $\tau \geq \left(\frac{\bar{\phi}_k}{\int_{\bar{\Xi}} \mathbf{x}^T \mathbf{\Lambda_w} \mathbf{x}\, \mathbb{m}(d\mathbf{x})}\right)^{\frac{\alpha}{2}} \geq \lambda \left(\frac{\lambda \bar{\phi}_k}{\mathbf{Tr}(\mathbf{Y} \mathbf{\Lambda_w})}\right)^{\frac{\alpha}{2}}$. The first inequality comes from Jensen's inequality and the second inequality is due to the fact that $\int_{\bar{\Xi}} \mathbf{x}\mathbf{x}^T \mathbb{m}(d\mathbf{x}) = \mathbf{Y}$ and $\lambda \leq 1$. Lastly, we verify that the same objective value can be achieved by the construction in (32): $\mathbb{E}_{\mathbb{m}}[p_k^\alpha(\mathbf{x})] = \int_{\bar{\Xi}} \mathbb{m}(d\mathbf{x}) - \int_{\bar{\Xi}} \big(1 - p_k^\alpha(\mathbf{x})\big) \mathbb{m}(d\mathbf{x}) = \lambda - \tau$. This completes the proof.

B. Proof of Proposition 6

It is easy to see that (16a) is equivalent to (15a) . Thus, we only need to show the equivalence between (15b) and (16b)-(16c). By the construction (17) and the inequalities in (16b)-(16c), all power exponents in (15b) are non-negative. Thus, we have

$$\tau^2\lambda^{\alpha-2}s^{2^l-\alpha} = \prod_{0\le m\le 2^l-1} t_{2^l+m} \ge \left(\prod_{0\le m\le 2^{l-1}-1} t_{2^{l-1}+m}\right)^2 \ge \cdots \ge t_1^{2^l} \ge s^{2^l}.$$

The converse is also true. Indeed, for any (τ,λ,s) that is feasible for (15b), we can construct a set of values $\{t_{2^{l-1}+m}\}$ such that $t_{2^l+2m}t_{2^l+2m+1} = t_{2^{l-1}+m}^2$ and $\tau^2\lambda^{\alpha-2}s^{2^l-\alpha} = \prod_{0\le m\le 2^{l-1}-1} t_{2^{l-1}+m}^2 \ge s^{2^l}$. The claim then follows by unrolling the above inequality.

C. Proof of Proposition 7

For any $\alpha \ge 2$ and $\mathbf{\Gamma} = (\mathbf{Y},\mathbf{y},\lambda,\tau)$ that is feasible for (14), let $u_b(\alpha,\mathbf{\Gamma}) = \lambda - \tau$ denote the objective value of $\mathbf{\Gamma}$. Now, consider a fixed $\alpha \ge 2$ and let $\mathbf{\Gamma}^\star = (\mathbf{Y}^\star,\mathbf{y}^\star,\lambda^\star,\tau^\star)$ be an optimal solution to (14). Note that the inequality constraint (14d) is satisfied as an equality at $\mathbf{\Gamma}^\star$; i.e., $\tau^\star = \lambda^\star\left(\frac{\lambda^\star\bar{\phi}}{\mathbf{Tr}(\mathbf{\Lambda_w}\mathbf{Y}^\star)}\right)^{\frac{\alpha}{2}}$. Hence, we have $\frac{\mathbf{Tr}(\mathbf{\Lambda_w}\mathbf{Y}^\star)}{\lambda^\star\bar{\phi}} = \left(\frac{\lambda^\star}{\tau^\star}\right)^{\frac{2}{\alpha}} \ge 1$, which implies that

$$\frac{\partial u_b(\alpha,\mathbf{\Gamma}^\star)}{\partial\alpha} = \frac{\lambda^\star}{2}\left(\frac{\lambda^\star\bar{\phi}}{\mathbf{Tr}(\mathbf{\Lambda_w}\mathbf{Y}^\star)}\right)^{\frac{\alpha}{2}}\log\left(\frac{\mathbf{Tr}(\mathbf{\Lambda_w}\mathbf{Y}^\star)}{\lambda^\star\bar{\phi}}\right) \ge 0.$$

In particular, for any $0 < \alpha_1 \le \alpha_2$, we have $u_b(\alpha_1,\mathbf{\Gamma}_1^\star) \le u_b(\alpha_2,\mathbf{\Gamma}_1^\star) \le u_b(\alpha_2,\mathbf{\Gamma}_2^\star)$, where the first inequality is due to the monotonicity of $u_b(\alpha,\mathbf{\Gamma}_1^\star)$ with respect to α and the second inequality is due to the fact that $\mathbf{\Gamma}_2^\star$ is an optimal solution corresponding to α_2. This completes the proof.

REFERENCES

[1] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surv. Tut.*, vol. 16, no. 4, pp. 1801–1819, 2014.

[2] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, S. Li, and G. Feng, "Device-to-device communications in cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 49–55, Apr. 2014.

[3] Y. Cao, T. Jiang, and C. Wang, "Cooperative device-to-device communications in cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 124–129, Jun. 2015.

[4] T. Kim and M. Dong, "An iterative Hungarian method to joint relay selection and resource allocation for D2D communications," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 625–628, Dec. 2014.

[5] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication: A message passing approach," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6326–6341, Nov. 2014.

[6] M. Fadel, A. El-Keyi, and A. Sultan, "QOS-constrained multiuser peer-to-peer amplify-and-forward relay beamforming," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1397–1408, Mar. 2012.

[7] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.

[8] S. H. Seyedmehdi and G. Boudreau, "An efficient clustering algorithm for device-to-device assisted virtual MIMO," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1334–1343, Mar. 2014.

[9] B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum sharing scheme between cellular users and ad-hoc device-to-device users," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1038–1049, Mar. 2013.

[10] Y. Xu, X. Zhao, and Y. C. Liang, "Robust power control and beamforming in cognitive radio networks: A survey," *IEEE Commun. Surv. Tut.*, vol. 17, no. 4, pp. 1834–1857, Fourthquarter 2015.

[11] M. Lin, J. Ouyang, and W. P. Zhu, "Joint beamforming and power control for device-to-device communications underlaying cellular networks," *IEEE J. Sel. Area. Commun.*, vol. 34, no. 1, pp. 138–150, Jan. 2016.

[12] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1668–1683, Apr. 2016.

[13] X. He, "Quadratically perturbed chance constrained programming with fitted distribution: t-distribution vs. Gaussian," 2015. [Online]. Available: http://www.optimization-online.org/DB_HTML/2015/05/4936.html

[14] W. W.-L. Li, Y. J. Zhang, A. M.-C. So, and M. Z. Win, "Slow adaptive OFDMA systems through chance constrained programming," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3858–3869, 2010.

[15] Y. J. Zhang and A. M.-C. So, "Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 362–373, 2011.

[16] Y. Qin, M. Ding, M. Zhang, H. Yu, and H. Luo, "Relaying robust beamforming for device-to-device communication with channel uncertainty," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1859–1862, Oct. 2014.

[17] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[18] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication under channel uncertainties: A stable matching approach," *IEEE Trans. Wireless Commun.*, vol. 63, no. 10, pp. 3882–3897, Oct. 2015.

[19] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056–3070, Jun. 2013.

[20] I. Pólik and T. Terlaky, "A survey of the S-lemma," *SIAM Review*, vol. 49, no. 3, pp. 371–418, 2007.

[21] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[22] Q. Li, A. M.-C. So, and W.-K. Ma, "Distributionally robust chance-constrained transmit beamforming for multiuser MISO downlink," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'14)*, May. 2014, pp. 3479–3483.

[23] L. Vandenberghe, S. Boyd, and K. Comanor, "Generalized Chebyshev bounds via semidefinite programming," *SIAM Review*, vol. 49, 2007.

[24] S. Gong, P. Wang, and L. Duan, "Distributed power control with robust protection for PUs in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3247–3258, Jun. 2015.

[25] S. X. Wu, X. Ni, and A. M.-C. So, "A polynomial optimization approach for the robust beamforming design in a device-to-device two-hop one-way relay network," in *proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16)*, Mar. 2016.

[26] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Proces. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[27] B. P. V. Parys, P. J. Goulart, and D. Kuhn, "Generalized Gauss inequalities via semidefinite programming," *Mathematical Programming*, pp. 1–32, 2015.

[28] S. Gong, S. X. Wu, A. M.-C. So, and X. Huang, "Distributionally robust relay beamforming in wireless communications," in *proc. ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'16)*, Malta, Nov. 2016.

[29] F. Gao, R. Zhang, and Y.-C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.

[30] L. V. J. Salo and P. Vainikainen, "Why is shadow fading lognormal?" in *proc. International Symposium on Wireless Personal Multimedia Communications (WPMC'05)*, Sep. 2005, pp. 522–526.

[31] S. Loyka, V. Kostina, and F. Gagnon, "On convexity of error rates in digital communications," *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6501–6516, Oct. 2013.

[32] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[33] M. V. Khlebnikov and P. S. Shcherbakov, "Petersen's lemma on matrix uncertainty and its generalizations," *Automation and Remote Control*, vol. 69, no. 11, pp. 1932–1945, 2008.

[34] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 664–678, Jan. 2010.

[35] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel Sci. Comput.*, vol. 11, no. 4, pp. 351–368, Dec. 2003.