A Distributionally Robust Minimum Variance Beamformer Design

Bin Li, Yue Rong, *Senior Member, IEEE*, Jie Sun, and Kok Lay Teo, *Senior Member, IEEE*

***Abstract*—This letter is concerned with a robust minimum variance beamformer design. To hedge the mismatch between the true and the assumed steering vectors, a distributionally robust beamformer (DR-beamformer) is proposed. The tractable reformulation of this beamformer is developed. Compared with the existing robust beamformers (e.g., worst-case robust beamformer and Gaussian robust beamformer), the proposed robust beamformer does not assume full knowledge of the channel mismatch. Therefore, it is more flexible in practice and more general in formulation. In addition, the relationships of the proposed robust beamformer to the existing ones are investigated. The performance gain of the DR-beamformer over the other robust beamformers is highlighted through numerical simulations.**

***Index Terms*—Distributionally robust optimization, gaussian robust beamformer, robust minimum variance beamformer (RMVB), worst-case robust beamformer.**

I. Introduction

MINIMUM variance beamformer (MVB) is an attractive adaptive beamforming approach. The optimal MVB can be obtained by using the well-known Capon beamformer [1]. However, the optimal MVB is designed based on the assumption that the steering vector is perfectly known. To combat the mismatch between the real and the assumed steering vectors, robust minimum variance beamformer (RMVB) design has been intensively studied [2]–[10]. These RMVBs can be classified into two main categories—the worst-case-based robust beamformers [2]–[7] and the chance-constraint-based robust beamformers [8]–[10].

A more general RMVB, distributionally robust beamformer (DR-beamformer), is proposed in this letter. The proposed RMVB provides a more general and practical formulation, in which we do not know the distribution of the steering vector mismatch. The term distributionally robust is introduced from the concept in distributionally robust optimization [11]–[21]. The proposed DR-beamformer is shown to have a better performance than the worst-case robust beamformers, the Gaussian robust beamformers, and other existing RMVBs.

Manuscript received August 7, 2017; revised October 24, 2017; accepted October 30, 2017. Date of publication November 15, 2017; date of current version December 7, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61701124 and in part by the Australian Research Council under Grant DP160102829. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Augusto Aubry. *(Corresponding author: Bin Li.)*

B. Li is with the School of Electrical Engineering and Information, Sichuan University, Chengdu 610000, China (e-mail: bin.li@scu.edu.cn).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Perth, W.A. 6102, Australia (e-mail: y.rong@curtin.edu.au).

J. Sun and K. L. Teo are with the Department of Mathematics and Statistics, Curtin University, Perth, W.A. 6102, Australia (e-mail: jie.sun@curtin.edu.au; k.l.teo@curtin.edu.au).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2773601

II. System Model and Backgrounds

A. System Model

We consider a sensor array that consists of M elements. A narrow-band signal $s(k)$, where k is the sample index, arrives at this array with a steering vector $\boldsymbol{a}$. Let $\boldsymbol{i}(k) \in \mathbb{C}^M$ and $\boldsymbol{n}(k) \in \mathbb{C}^M$ be the interference vector and the noise vector at the kth sample, respectively. Then, the observation or the output of the sensor array is given by

$$\boldsymbol{x}(k) = s(k)\boldsymbol{a} + \boldsymbol{i}(k) + \boldsymbol{n}(k). \tag{1}$$

A narrow-band adaptive beamformer is applied at the sensor array and its output is given by

$$y(k) = \boldsymbol{w}^H \boldsymbol{x}(k) \tag{2}$$

where $\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^T$ is the complex vector of beamformer weights, and $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and Hermitian transpose, respectively. By substituting (1) into (2), it follows that

$$y(k) = \boldsymbol{w}^H \boldsymbol{a} s(k) + \boldsymbol{w}^H \boldsymbol{i}(k) + \boldsymbol{w}^H \boldsymbol{n}(k). \tag{3}$$

The performance of a beamformer is usually measured by the output signal-to-interference-plus-noise ratio (SINR), which is given by

$$\text{SINR} = \frac{\sigma_s^2 \left|\boldsymbol{w}^H \boldsymbol{a}\right|^2}{\boldsymbol{w}^H \boldsymbol{R}_{i+n} \boldsymbol{w}} \tag{4}$$

where σ_s^2 is the power of the signal of interest, $|\cdot|$ denotes the absolute value, and $\boldsymbol{R}_{i+n} \in \mathbb{C}^{M \times M}$ is the interference-plus-noise covariance matrix defined as

$$\boldsymbol{R}_{i+n} = \mathbb{E}\left\{(\boldsymbol{i}(k) + \boldsymbol{n}(k))(\boldsymbol{i}(k) + \boldsymbol{n}(k))^H\right\}.$$

Here, $\mathbb{E}\{\cdot\}$ denotes the statistical expectation. However, $\boldsymbol{R}_{i+n}$ is usually not available in real-world applications. Instead, the sample covariance matrix

$$\hat{\boldsymbol{R}}_y = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}(k)\boldsymbol{x}^H(k) \tag{5}$$

is used in practice as an alternative of $\boldsymbol{R}_{i+n}$, where K is the size of the training sample. Note that advanced covariance matrix estimators [22] can also be used here, and $\boldsymbol{x}(k)$ here is not signal free in general.

The MVB is formulated to maximize the SINR by minimizing the output power of the sensor array while keeping a distortionless response of the desired signal, which is given by

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^H \hat{\boldsymbol{R}}_y \boldsymbol{w} \qquad \text{s.t.} \ \ \boldsymbol{w}^H \boldsymbol{a} = 1. \tag{6}$$

To hedge the mismatch between the real steering vector and the assumed one, we consider the RMVB in this letter. For the RMVB, the equality constraint in (6) is usually modified into an inequality constraint, see [3] and [8]. This is because the equality constraint cannot hold for all the possible steering vectors.

In this letter, we adopt the formulation as that in [3], and the MVB problem can be formulated as follows:

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^H \hat{\boldsymbol{R}}_y \boldsymbol{w} \qquad \text{s.t.} \ \ \mathrm{Re}\{\boldsymbol{w}^H \boldsymbol{a}\} \geq 1 \tag{7}$$

where $\mathrm{Re}\{\cdot\}$ stands for the real part of a complex number. This formulation provides an efficient lower bound for the magnitude of the response, which is particularly true when the uncertainty in the array response is relatively small. Then, we introduce the error vector

$$\boldsymbol{\delta} \triangleq \boldsymbol{a} - \tilde{\boldsymbol{a}} \tag{8}$$

where $\tilde{\boldsymbol{a}}$ and $\boldsymbol{\delta}$ denote the assumed steering vector and the mismatch vector, respectively. Using (8), problem (7) becomes

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^H \hat{\boldsymbol{R}}_y \boldsymbol{w} \qquad \text{s.t.} \ \mathrm{Re}\left\{\boldsymbol{w}^H (\tilde{\boldsymbol{a}} + \boldsymbol{\delta})\right\} \geq 1. \tag{9}$$

In order to transform problem (9) into a real-valued form, we define

$$\underline{\boldsymbol{w}} \triangleq \begin{bmatrix} \mathrm{Re}\{\boldsymbol{w}\} \\ \mathrm{Im}\{\boldsymbol{w}\} \end{bmatrix}, \quad \underline{\tilde{\boldsymbol{a}}} \triangleq \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{a}}\} \\ \mathrm{Im}\{\tilde{\boldsymbol{a}}\} \end{bmatrix}, \quad \underline{\boldsymbol{\delta}} \triangleq \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\delta}\} \\ \mathrm{Im}\{\boldsymbol{\delta}\} \end{bmatrix} \tag{10}$$

$$\boldsymbol{R} \triangleq \begin{bmatrix} \mathrm{Re}\left\{\hat{\boldsymbol{R}}_{\mathrm{y}}\right\} & -\mathrm{Im}\left\{\hat{\boldsymbol{R}}_{\mathrm{y}}\right\} \\ \mathrm{Im}\left\{\hat{\boldsymbol{R}}_{\mathrm{y}}\right\} & \mathrm{Re}\left\{\hat{\boldsymbol{R}}_{\mathrm{y}}\right\} \end{bmatrix}. \tag{11}$$

By considering (10) and (11), problem (9) can be rewritten into the following real-valued form:

$$\min_{\underline{\boldsymbol{w}}} \ \underline{\boldsymbol{w}}^T \boldsymbol{R} \underline{\boldsymbol{w}} \qquad \text{s.t.} \ \underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1. \tag{12}$$

B. Worst-Case Robust and the Gaussian Robust Beamformers

The worst-case-based beamformer [3] is formulated as

$$\min_{\underline{\boldsymbol{w}}} \underline{\boldsymbol{w}}^T \boldsymbol{R} \underline{\boldsymbol{w}} \tag{13}$$

$$\text{s.t.} \min_{\|\boldsymbol{\delta}\| \leq \eta} \underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1 \tag{14}$$

where $\|\cdot\|$ denotes the Euclidian norm. According to [3], the second-order cone programming (SOCP) reformulation of problems (13) and (14) is

$$\min_{\underline{\boldsymbol{w}}, \tau} \ \tau \tag{15}$$

$$\text{s.t.} \ \ \|\boldsymbol{L}\underline{\boldsymbol{w}}\| \leq \tau \tag{16}$$

$$\eta \|\underline{\boldsymbol{w}}\| \leq \underline{\boldsymbol{w}}^T \underline{\tilde{\boldsymbol{a}}} - 1 \tag{17}$$

where $\boldsymbol{R} = \boldsymbol{L}^T \boldsymbol{L}$.

The Gaussian robust beamformer [8] design is formulated as

$$\min_{\underline{\boldsymbol{w}}} \ \underline{\boldsymbol{w}}^T \boldsymbol{R} \underline{\boldsymbol{w}} \tag{18}$$

$$\text{s.t.} \ \ \mathrm{Pr}_{[\mathbb{G}]} \left[\underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1\right] \geq p \tag{19}$$

where $\mathrm{Pr}_{[\mathbb{G}]}[\cdot]$ stands for the probability under the Gaussian distribution $\mathbb{G}$, and $p \in (0, 1)$ denotes the probability that the inequality $\underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1$ holds. Here, $\boldsymbol{\delta}$ is subject to Gaussian distribution. The SOCP reformulation of problems (18) and (19) can be obtained similar to that in [8]

$$\min_{\underline{\boldsymbol{w}}, \tau} \ \tau \tag{20}$$

$$\text{s.t.} \ \ \|\boldsymbol{L}\underline{\boldsymbol{w}}\| \leq \tau \tag{21}$$

$$\sigma \mathrm{erf}^{-1}(2p - 1) \|\underline{\boldsymbol{w}}\| \leq \underline{\boldsymbol{w}}^T \underline{\tilde{\boldsymbol{a}}} - 1 \tag{22}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t \tag{23}$$

denotes the standard error function for the Gaussian distribution.

III. Distributionally Robust Beamformer

In this section, we develop a more general robust beamformer—DR-beamformer. More specifically, the beamformer design is based only on the first-order and second-order moments of the channel mismatch. Here, we assume that the stochastic distribution of $\boldsymbol{\delta}$, which is denoted as $\mathbb{P}$, is within the following set:

$$\mathcal{P} = \left\{\mathbb{P} \,\middle|\, \mathbb{E}_{[\mathbb{P}]}\{\boldsymbol{\delta}\} = \boldsymbol{0}, \ \mathbb{E}_{[\mathbb{P}]}\{\boldsymbol{\delta}\boldsymbol{\delta}^T\} = \sigma^2 \boldsymbol{I}\right\}. \tag{24}$$

The robustness here is in the sense of finding the worst-case distribution among all the possible distributions in $\mathcal{P}$ rather than the Gaussian distribution such that the chance constraint

$$\mathrm{Pr}_{[\mathbb{P}]} \left[\underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1\right] \geq p$$

is satisfied. Therefore, we formulate the DR-beamformer as follows:

$$\min_{\underline{\boldsymbol{w}}} \ \underline{\boldsymbol{w}}^T \boldsymbol{R} \underline{\boldsymbol{w}} \tag{25}$$

$$\text{s.t.} \ \inf_{\mathbb{P} \in \mathcal{P}} \mathrm{Pr}_{[\mathbb{P}]} \left[\underline{\boldsymbol{w}}^T (\underline{\tilde{\boldsymbol{a}}} + \underline{\boldsymbol{\delta}}) \geq 1\right] \geq p. \tag{26}$$

Without the Gaussian distribution assumption, the chance constraint (26) is usually nonconvex [23] and hence it is difficult to solve problems (25) and (26). In the following, we derive a tractable reformulation of problems (25) and (26).

Theorem 1: Problems (25) and (26) can be reformulated as the following conic optimization problem:

$$\min_{\tau, \beta, \underline{\boldsymbol{w}}, \boldsymbol{M}} \ \tau \tag{27}$$

$$\text{s.t.} \ \ \|\boldsymbol{L}\underline{\boldsymbol{w}}\| \leq \tau \tag{28}$$

$$\beta + \frac{1}{1-p} \mathrm{Tr}(\boldsymbol{\Omega}\boldsymbol{M}) \leq 0 \tag{29}$$

$$\boldsymbol{M} + \begin{bmatrix} \boldsymbol{0} & \frac{1}{2}\underline{\boldsymbol{w}} \\ \frac{1}{2}\underline{\boldsymbol{w}}^T & \underline{\boldsymbol{w}}^T \underline{\tilde{\boldsymbol{a}}} + \beta - 1 \end{bmatrix} \succeq 0 \tag{30}$$

$$\boldsymbol{M} \succeq 0 \tag{31}$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace

$$\boldsymbol{\Omega} = \begin{bmatrix} \frac{\sigma^2}{2}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \tag{32}$$

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISONS

RMVB	Worst-case	Gaussian	DR-beamformer
Complexity	$\mathcal{O}(M^3T^3)$	$\mathcal{O}(M^3T^3)$	$\mathcal{O}(M^{4.5}T^{4.5})$

and $\boldsymbol{M} \in \mathbb{S}^{2MN+1}$ is the set consisting of all the $(2MN+1) \times (2MN+1)$ symmetric matrices.

Proof: See the Appendix. ■

To establish the relationships between the proposed DR-beamformer and the other RMVBs, we rewrite $\boldsymbol{M}$ as

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{m}_2 \\ \boldsymbol{m}_2^T & m_3 \end{bmatrix}. \tag{33}$$

From (29) and (33), we have

$$-\beta \geq \frac{1}{1-p}\mathrm{Tr}\left(\boldsymbol{\Omega M}\right) = \frac{1}{1-p}\left(\frac{\sigma^2}{2}\mathrm{Tr}\left(\boldsymbol{M}_1\right) + m_3\right). \tag{34}$$

By considering (30) and (33), we have

$$\begin{bmatrix} \boldsymbol{M}_1 & \frac{1}{2}\underline{\boldsymbol{w}} + \boldsymbol{m}_2 \\ \frac{1}{2}\underline{\boldsymbol{w}}^T + \boldsymbol{m}_2^T & \underline{\boldsymbol{w}}^T\underline{\tilde{\boldsymbol{a}}} + \beta - 1 + m_3 \end{bmatrix} \succeq 0. \tag{35}$$

Then, we apply Schur-complement [25] to (35) yielding

$$\underline{\boldsymbol{w}}^T\underline{\tilde{\boldsymbol{a}}} + \beta - 1 + m_3 - \left(\frac{1}{2}\underline{\boldsymbol{w}}^T + \boldsymbol{m}_2^T\right)\boldsymbol{M}_1^{\dagger}\left(\frac{1}{2}\underline{\boldsymbol{w}} + \boldsymbol{m}_2\right) \geq 0 \tag{36}$$

where $(\cdot)^{\dagger}$ is matrix pseudo-inverse. Together with (34), we have

$$\begin{aligned} &\underline{\boldsymbol{w}}^T\underline{\tilde{\boldsymbol{a}}} - 1 \\ &\geq -\beta + \left(\frac{1}{2}\underline{\boldsymbol{w}}^T + \boldsymbol{m}_2^T\right)\boldsymbol{M}_1^{\dagger}\left(\frac{1}{2}\underline{\boldsymbol{w}} + \boldsymbol{m}_2\right) - m_3 \\ &\geq \frac{1}{1-p}\left(\frac{\sigma^2}{2}\mathrm{Tr}\left(\boldsymbol{M}_1\right) + m_3\right) - m_3 \\ &\quad + \left(\frac{1}{2}\underline{\boldsymbol{w}}^T + \boldsymbol{m}_2^T\right)\boldsymbol{M}_1^{\dagger}\left(\frac{1}{2}\underline{\boldsymbol{w}} + \boldsymbol{m}_2\right) \\ &= \frac{1}{4}\underline{\boldsymbol{w}}^T\boldsymbol{M}_1^{\dagger}\underline{\boldsymbol{w}} + c \end{aligned} \tag{37}$$

where $c = \frac{\sigma^2}{2(1-p)}\mathrm{Tr}\left(\boldsymbol{M}_1\right) + \frac{p}{1-p}m_3 + \boldsymbol{m}_2^T\boldsymbol{M}_1^{\dagger}\boldsymbol{m}_2 + \frac{1}{2}(\underline{\boldsymbol{w}}^T\boldsymbol{M}_1^{\dagger}\boldsymbol{m}_2 + \boldsymbol{m}_2^T\boldsymbol{M}_1^{\dagger}\underline{\boldsymbol{w}})$.

By comparing (37) with (17) and (22), we can see that when $\boldsymbol{M}_1^{\dagger} = \kappa\boldsymbol{I}$, (37) becomes (17) if $\kappa = 4\eta$ and $c = 0$, while (37) becomes (22) if $\kappa = 4\sigma\mathrm{erf}^{-1}(2p-1)$ and $c = 0$. Thus, our design is more general.

According to the results in [26], we list the computational complexity order of three beamformers in Table I. It can be seen from Table I that the DR-beamformer has a higher computational complexity order than the other two robust beamformers. It will be seen in the next section that the proposed DR-beamformer has a better SINR performance than the other two robust beamformers. Indeed, the performance-complexity tradeoff among these RMVBs is interesting for real-world applications.

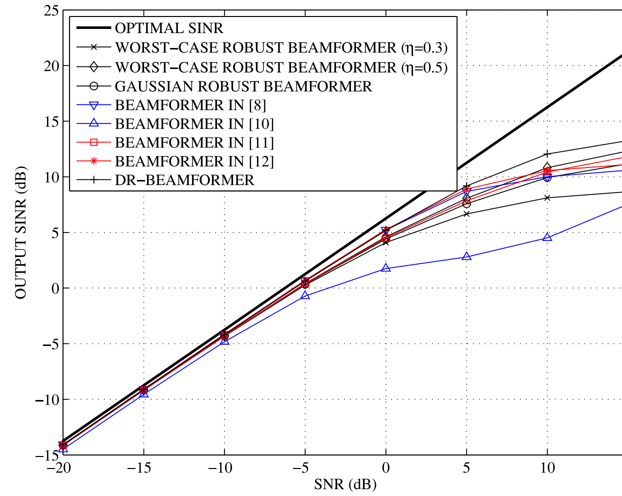

Fig. 1. Example 1: Gaussian mixture mismatch.

IV. SIMULATIONS

We consider a uniform linear array with $M = 5$ omnidirectional sensor elements, and each element is spaced half a wavelength apart from the adjacent ones. The number of training samples is set as $K = 80$. There are two interfering sources with plane waveforms, and their directions of arrivals are 20° and 50°. The interference-to-noise ratio is set as 20 dB. We fix the assumed steering vector $\tilde{\boldsymbol{a}} = \left[1, e^{j\pi\sin(\theta)}, \ldots, e^{j\pi\sin((M-1)\theta)}\right]^T$ with $\theta = 5°$ (0.087 radian) throughout the simulations. The CVX toolbox [24] is used to compute the weight vectors of the worst-case robust beamformer (13), (14), the Gaussian robust beamformer (18), (19), the robust beamformers in [11] and [12], [8, Section III-B], [10, Section IV], and the DR-beamformer (27)–(31). The mismatch vector $\boldsymbol{\delta}$ in (8) is generated randomly based on distributions detailed later. For each realization of $\boldsymbol{a}$, the optimal SINR and the output SINR of each beamformer are evaluated with the actual $\boldsymbol{a}$ and the true $\boldsymbol{R}_{i+n}$ as (4). All simulation results are averaged through 1000 realizations of $\boldsymbol{a}$.

In the first example, we choose $\eta = 0.3$ and 0.5 in (14), and p is set as 0.9. We consider the Gaussian mixture model, which is widely used to approximate the non-Gaussian noise in communications [27]. The probability density function (pdf) of δ_m is given as

$$f(\delta_m) = \sum_{l=1}^{L}\frac{\lambda_l}{\pi\sigma_l^2}\exp\left\{-\frac{|\delta_m|^2}{\sigma_l^2}\right\}, \quad m = 1, \ldots, M \tag{38}$$

where $\sum_{l=1}^{L}\lambda_l = 1$. According to [27], (38) is a spherically symmetric bivariate pdf for the complex-valued random variable δ_m. In particular, for the case of $L = 2$, it is a typical model for impulsive noise if $\sigma_2 \gg \sigma_1$ and $\lambda_2 < \lambda_1$. Fig. 1 shows the beamformer output SINR versus the signal-to-noise ratio (SNR) for a steering vector mismatch scenario with $\sigma_1^2 = 0.1, \lambda_1 = 0.9$, $\sigma_2^2 = 1.2$, and $\lambda_2 = 0.1$. It can be seen from Fig. 1 that the proposed DR-beamformer yields the highest output SINR among all eight beamformers tested.

In the second example, a Gaussian steering vector mismatch is considered with zero-mean and variance of 0.1. We set $\eta = 0.3$ and 0.5, and $p = 0.95$. We plot the output SINR versus the SNR in Fig. 2. It can be seen from Fig. 2 that the performance of the DR-beamformer is better than the other seven robust beamformers. Note that because the mismatch vector $\boldsymbol{\delta}$ has

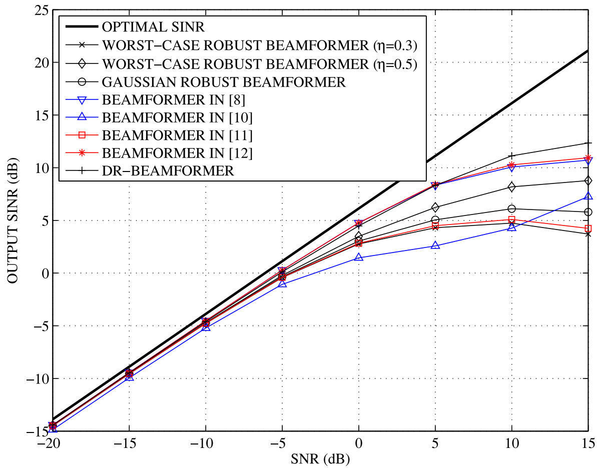

Fig. 2. Example 2: Gaussian mismatch.

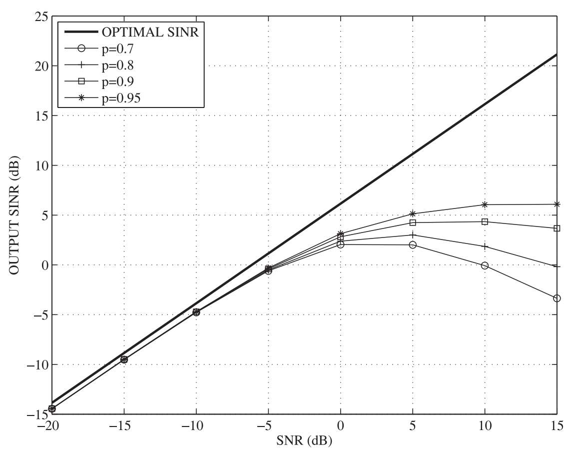

Fig. 3. Example 2: Gaussian robust beamformer with various p.

different distributions in two simulation examples, the behavior of some curves in Fig. 2 is slightly different to that in Fig. 1.

Interestingly, it can be observed from Fig. 2 that although $\boldsymbol{\delta}$ is Gaussian, the proposed beamformer has a better performance than the Gaussian robust beamformer. The reasons are explained below. Let us consider problems (25) and (26). As the constraint $\Pr_{[\mathbb{P}]}\left[\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})\geq 1\right]\geq p$ needs to hold for the worst-case distribution in the distribution space, there is $\Pr_{[\mathbb{G}]}\left[\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})\geq 1\right]\geq p_1$ with $p_1\geq p$. Therefore, when $\boldsymbol{\delta}$ has a Gaussian distribution, problems (25) and (26) become

$$\min_{\underline{\boldsymbol{w}}}\ \underline{\boldsymbol{w}}^T\boldsymbol{R}\underline{\boldsymbol{w}}\qquad \text{s.t.}\ \Pr_{[\mathbb{G}]}\left[\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})\geq 1\right]\geq p_1 \tag{39}$$

which is essentially the Gaussian robust beamformer (18), (19) with $p_1\geq p$. It can be seen from Fig. 3 (with same simulation setups as shown in Fig. 2) that for the Gaussian robust beamformer, the output SINR increases with p until p approaches to 1 (where the problem becomes infeasible). Therefore, the proposed DR-beamformer has a better SINR performance than the Gaussian robust beamformer for $p<1$.

V. Conclusion

A more general distributionally robust optimization based beamformer is formulated, and the corresponding tractable reformulation is derived. The computational complexities of the RMVBs are compared. Particularly, a better performance of the DR-robust beamformer has been demonstrated in terms of its general formulation.

Appendix
Proof of Theorem 1

The following results are needed to prove Theorem 1.

Lemma 1: [18, Theorem 2.2]: Let $L:\mathbb{R}^k\rightarrow\mathbb{R}$ be a continuous loss function that is either concave in $\tilde{\boldsymbol{\xi}}$ or quadratic in $\tilde{\boldsymbol{\xi}}$. Then, the following equivalence holds:

$$\inf_{\mathbb{P}\in\mathcal{P}}\Pr_{[\mathbb{P}]}\left[L(\tilde{\boldsymbol{\xi}})\leq 0\right]\geq 1-\epsilon \iff \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}-\mathrm{CVaR}_\epsilon\left[L(\tilde{\boldsymbol{\xi}})\right]\leq 0 \tag{40}$$

where $\tilde{\boldsymbol{\xi}}$ is a random vector and $\mathcal{P}$ is defined in (24), and

$$\mathbb{P}-\mathrm{CVaR}_\epsilon\left[\tilde{\xi}\right]=\inf_{\beta\in\mathbb{R}}\left\{\beta+\frac{1}{\epsilon}\mathbb{E}_{[\mathbb{P}]}\left[\left(\tilde{\xi}-\beta\right)^+\right]\right\}$$

where $\mathbb{R}$ denotes the set of all real numbers and $(a)^+=\max(a,0)$ for a real number a.

Lemma 2: [18, Theorem 21]: The feasible set

$$\left\{\boldsymbol{x}\in\mathbb{R}^n:\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}-\mathrm{CVaR}_\epsilon\left[y^0(\boldsymbol{x})+\boldsymbol{y}^T(\boldsymbol{x})\tilde{\boldsymbol{\xi}}\right]\leq 0\right\} \tag{41}$$

can be written as

$$\left\{\boldsymbol{x}\in\mathbb{R}^n:\begin{array}{c}\boldsymbol{M}\succeq 0,\ \beta+\frac{1}{\epsilon}\mathrm{Tr}(\boldsymbol{\Omega}\boldsymbol{M})\leq 0,\\ \boldsymbol{M}-\begin{bmatrix}\boldsymbol{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})\\ \frac{1}{2}\boldsymbol{y}^T(\boldsymbol{x}) & y^0(\boldsymbol{x})-\beta\end{bmatrix}\succeq 0\end{array}\right\} \tag{42}$$

where

$$\boldsymbol{\Omega}=\begin{bmatrix}\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^T & \boldsymbol{\mu}\\ \boldsymbol{\mu}^T & 1\end{bmatrix} \tag{43}$$

$\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance of $\tilde{\boldsymbol{\xi}}$, $y^0(\boldsymbol{x})$ and $\boldsymbol{y}(\boldsymbol{x})$ depend affinely on $\boldsymbol{x}$.

Now, we start to prove Theorem 1. In view of (26), we can see that $L(\underline{\boldsymbol{\delta}})=1-\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})$ depends affinely on $\underline{\boldsymbol{\delta}}$, which also implies that $L(\underline{\boldsymbol{\delta}})$ is concave in $\underline{\boldsymbol{\delta}}$. Thus, from Lemma 1, we know that (26) is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}-\mathrm{CVaR}_{1-p}\left[1-\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})\right]\leq 0 \tag{44}$$

where

$$\mathbb{P}-\mathrm{CVaR}_{1-p}\left[1-\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})\right]=\inf_{\beta\in\mathbb{R}}\left\{\beta+\frac{1}{1-p}\mathbb{E}_{[\mathbb{P}]}\left[\left(1-\underline{\boldsymbol{w}}^T(\underline{\tilde{\boldsymbol{a}}}+\underline{\boldsymbol{\delta}})-\beta\right)^+\right]\right\}.$$

Let $y^0(\underline{\boldsymbol{w}})=1-\underline{\boldsymbol{w}}^T\underline{\tilde{\boldsymbol{a}}}$ and $\boldsymbol{y}(\boldsymbol{w}_k)=-\underline{\boldsymbol{w}}$. Clearly, $y^0(\underline{\boldsymbol{w}})$ and $\boldsymbol{y}(\underline{\boldsymbol{w}})$ affinely depend on $\underline{\boldsymbol{w}}$. Here, $\boldsymbol{\mu}=\boldsymbol{0}$ and $\boldsymbol{\Sigma}=\frac{\sigma^2}{2}\boldsymbol{I}$. Then, from Lemma 2, (44) can be rewritten as constraints (29)–(31). Note that $\boldsymbol{R}=\boldsymbol{L}^T\boldsymbol{L}$. By introducing a new decision variable τ, (25) can be rewritten in the epigraph form as that in (27) and (28). This completes the proof. ■

REFERENCES

[1] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407–2423, Sep. 2004.

[2] S. Vorobyov, A. B. Gershman, Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[3] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[4] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 328–340, Mar. 1984.

[5] S.-J. Kim, A. Magnani, A. Mutapcic, S. P. Boyd, and Z.-Q. Luo, "Robust beamforming via worst-case SINR maximization," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 1539–1547, Apr. 2008.

[6] A. Beck and Y. Eldar, "Doubly constrained robust Capon beamformer with ellipsoidal uncertainty sets," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 753–758, Feb. 2007.

[7] A. Aubry, A. De Maio, Y. Huang, and M. Piezzo, "Robust design of radar Doppler filters," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5848–5860, Nov. 2016.

[8] S. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5719–5724, Nov. 2008.

[9] S. Vorobyov, Y. Rong, and A. B. Gershman, "Robust adaptive beamforming using probability-constrained optimization," in *Proc. 13th IEEE Workshop Statist. Signal Process.*, Bordeaux, Jul. 2005, pp. 934–939.

[10] W. Zhang and S. Vorobyov, "Joint robust transmit/receive adaptive beamforming for MIMO radar using probability-constrained optimization," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 112–116, Jan. 2016.

[11] X. Zhang, Y. Li, N. Ge, and J. Lu, "Robust minimum variance beamforming under distributional uncertainty," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, Apr. 2015, pp. 2514–2518.

[12] X. Zhang, Q. Feng, N. Ge, and J. Lu, "Distributionally robust chance-constrained minimum variance beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, Mar. 2016, pp. 2881–2885.

[13] Q. Li, A. M.-C. So, and W. K. Ma, "Distributionally robust chance-constrained transmit beamforming for multiuser MISO downlink," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, May 2014, pp. 3479–3483.

[14] A. M.-C. So and Y. J. Zhang, "Distributionally robust slow adaptive OFDMA with soft QoS via linear programming," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 947–958, May 2013.

[15] B. Li, Y. Rong, J. Sun, and K. L. Teo, "A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 464–474, Jan. 2017.

[16] X. He and Y. C. Wu, "Probabilistic QoS constrained robust downlink multiuser MIMO transceiver design with arbitrarily distributed channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6292–6302, Dec. 2013.

[17] X. He and Y. C. Wu, "Tight probabilistic SINR constrained beamforming under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3490–3505, Jul. 2015.

[18] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program. A*, vol. 137, pp. 167–198, 2013.

[19] G. Calafiore and L. El Ghaoui, "Distributionally robust chance constraints linear programs with applications," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1–22, 2006.

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

[21] A. M.-C. So, "Moment inequalities for sums of random matrices and their applications in optimization," *Math. Program.*, vol. 130, no. 1, pp. 125–151, 2011.

[22] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

[23] W. Chen, M. Sim, J. Sun, and C. P. Teo, "From CVaR to uncertainty set: Implications in joint-chance-constrained optimization," *Oper. Res.*, vol. 58, no. 2, pp. 470–485, 2010.

[24] M. Grant and S. P. Boyd, "CVX: Matlab software for disciplined convex programming," Sep. 2010. [Online]. Available: http://cvxr.com/cvx

[25] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] M. S. Lobo, L. Vandenberghe, and S. P. Boyd, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193–228, Nov. 1998.

[27] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.