DISTRIBUTIONALLY ROBUST OUTLIER-AWARE RECEIVE BEAMFORMING

Shixiong Wang, Wei Dai, and Geoffrey Ye Li

Department of Electrical & Electronic Engineering, Imperial College London, United Kingdom
(s.wang@u.nus.edu, wei.dai1@imperial.ac.uk, geoffrey.li@imperial.ac.uk)

ABSTRACT

This paper investigates the robust receive beamforming problem in wireless communication from the perspective of trustworthy machine learning. The inaccuracies of the signal transmission model, i.e., outliers in the received signals and channel uncertainties, are quantified by the distributional deviations from the nominal distribution induced by the nominal signal model. The worst-case signal estimation performance (i.e., worst-case mean-squared error) is minimized to achieve robustness against the above two types of distributional inaccuracies. The resultant receive beamformers incorporate clipping operations to mitigate the adverse effects caused by outliers. Experiments validate the effectiveness of the proposed robust receive beamforming method in suppressing outliers and combating channel uncertainties.

Index Terms— Distributional Robustness, Measurement Outlier, Model Uncertainty, Receive Beamforming

1. INTRODUCTION

In multi-input multi-output (MIMO) wireless communications, outliers may exist in received signals, which can be caused by, for example, array abnormality (e.g., manufacturing defects, physical damages) or impulse channel noises. Impulse channel noises may occur due to switching transients in power lines or automobile ignition in outdoor environments [1], to electromechanical switches of appliances in indoor surroundings [2], to multipath fading [3], to multiuser constructive interference, among many others. Ignoring the existence of outliers and directly employing outlier-unaware receive beamformers such as Wiener (i.e., minimum mean-squared error beamformer; MMSE), Capon (i.e., minimum variance distortionless response beamformer; MVDR), and Zeroforing may cause significant performance degradation in recovering the transmitted signals. For MIMO wireless communications, outlier-aware receive beamformers are seldom reported, although outlier-aware beamforming for wireless sensing has been intensively studied; see, e.g., [4–8]. In addition, conventional beamformers are also sensitive to various channel uncertainties in signal transmission [9], for example, the scarcity of pilot data. As an extension of [9] in response to outliers, this paper studies a distributionally robust outlier-aware (DROA) receive beamformer that can suppress the adverse influences introduced by outliers in the received signals and simultaneously combat the channel uncertainties in the signal transmission model. The method is derived from the perspective of trustworthy machine learning that combats the distributional uncertainty in the empirical distribution constructed using pilot data. The min-max formulation is leveraged where the minimization is to design the optimal beamformer, while the maximization is to find the worst-case distributional model uncertainty due to outliers and channel inaccuracies.

Notations: Random quantities are written in upright fonts, while deterministic ones are in italics. Matrices and vectors are denoted by uppercase and lowercase symbols, respectively. Let $\mathbb{C}^d$ denote the d-dimensional complex coordinate space. We use $\mathcal{CN}(\mathbf{0}, \boldsymbol{R}, \boldsymbol{C})$ to denote the complex Gaussian distribution with zero mean, covariance $\boldsymbol{R}$, and pseudo-covariance $\boldsymbol{C}$; when $\boldsymbol{C}$ is not specified, it is treated as a zero matrix. Let $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ denote the real Gaussian distribution with zero mean and covariance $\boldsymbol{\Sigma}$. The trace and inverse of square matrix $\boldsymbol{X}$ are denoted as $\operatorname{Tr} \boldsymbol{X}$ and $\boldsymbol{X}^{-1}$, respectively.

2. PROBLEM FORMULATION

We consider a base-band wireless signal transmission model

$$\mathbf{x} = \boldsymbol{H}\mathbf{s} + \mathbf{v}, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{s} \in \mathbb{C}^M$, $\boldsymbol{H} \in \mathbb{C}^{N\times M}$, and $\mathbf{v} \in \mathbb{C}^N$ denote the received signal, transmitted signal, channel matrix, channel noise, respectively. We suppose that $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R}_v)$. For technical convenience, this paper works on the real-space equivalent of (1) by stacking the real and imaginary components:

$$\underline{\mathbf{x}} = \underline{\boldsymbol{H}} \cdot \underline{\mathbf{s}} + \underline{\mathbf{v}}, \tag{2}$$

where

$$\underline{\mathbf{x}} := \begin{bmatrix} \operatorname{Re} \mathbf{x} \\ \operatorname{Im} \mathbf{x} \end{bmatrix}, \quad \underline{\mathbf{s}} := \begin{bmatrix} \operatorname{Re} \mathbf{s} \\ \operatorname{Im} \mathbf{s} \end{bmatrix}, \quad \underline{\mathbf{v}} := \begin{bmatrix} \operatorname{Re} \mathbf{v} \\ \operatorname{Im} \mathbf{v} \end{bmatrix},$$

This work is supported by the UK Department for Science, Innovation and Technology under the Future Open Networks Research Challenge project TUDOR (Towards Ubiquitous 3D Open Resilient Network). The views expressed are those of the authors and do not necessarily represent the project.

and

$$\underline{\boldsymbol{H}} := \left[\begin{array}{cc} \operatorname{Re}\boldsymbol{H} & -\operatorname{Im}\boldsymbol{H} \\ \operatorname{Im}\boldsymbol{H} & \operatorname{Re}\boldsymbol{H} \end{array}\right];$$

$\underline{\mathbf{x}} \in \mathbb{R}^{2N}$, $\underline{\mathbf{s}} \in \mathbb{R}^{2M}$, $\underline{\mathbf{v}} \in \mathbb{R}^{2N}$, and $\underline{\boldsymbol{H}} \in \mathbb{R}^{2N\times 2M}$.

We examine the receive beamforming problem from the perspective of statistical machine learning. To estimate the transmitted signal $\underline{\mathbf{s}}$, we solve the following least mean square problem

$$\min_{\boldsymbol{\phi}\in\mathcal{B}_{\mathbb{R}^{2N}\to\mathbb{R}^{2M}}} \operatorname{Tr}\mathbb{E}_{(\underline{\mathbf{x}},\underline{\mathbf{s}})\sim\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}}[\boldsymbol{\phi}(\underline{\mathbf{x}})-\underline{\mathbf{s}}][\boldsymbol{\phi}(\underline{\mathbf{x}})-\underline{\mathbf{s}}]^{\mathsf{T}}, \tag{3}$$

where an estimator $\boldsymbol{\phi}$ is a function from $\underline{\mathbf{x}}$ to $\underline{\mathbf{s}}$; $\mathcal{B}_{\mathbb{R}^{2N}\to\mathbb{R}^{2M}}$ contains all Borel measurable functions from $\mathbb{R}^{2N}$ to $\mathbb{R}^{2M}$; $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is the joint distribution of $(\underline{\mathbf{x}},\underline{\mathbf{s}})$. In what follows, we use $\mathcal{B}$ as a shorthand for $\mathcal{B}_{\mathbb{R}^{2N}\to\mathbb{R}^{2M}}$. The optimal estimator is known as the conditional mean of $\underline{\mathbf{s}}$ given $\underline{\mathbf{x}}$; i.e.,

$$\hat{\underline{\mathbf{s}}} = \boldsymbol{\phi}(\underline{\mathbf{x}}) = \mathbb{E}_{(\underline{\mathbf{x}},\underline{\mathbf{s}})\sim\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}}(\underline{\mathbf{s}}|\underline{\mathbf{x}}). \tag{4}$$

When $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is a joint Gaussian, the optimal estimator is of a linear form, i.e.,

$$\hat{\underline{\mathbf{s}}} = \boldsymbol{W}\underline{\mathbf{x}} = \boldsymbol{R}_{\underline{x}\underline{s}}^{\mathsf{T}}\boldsymbol{R}_{\underline{x}}^{-1}\underline{\mathbf{x}}, \tag{5}$$

where

$$\boldsymbol{W} := \boldsymbol{R}_{\underline{x}\underline{s}}^{\mathsf{T}}\boldsymbol{R}_{\underline{x}}^{-1} \tag{6}$$

is called the Wiener receive beamformer, $\boldsymbol{R}_{\underline{x}\underline{s}} := \mathbb{E}\underline{\mathbf{x}}\underline{\mathbf{s}}^{\mathsf{T}}$, and $\boldsymbol{R}_{\underline{x}} := \mathbb{E}\underline{\mathbf{x}}\underline{\mathbf{x}}^{\mathsf{T}}$. However, when outliers exist in the received signal $\underline{\mathbf{x}}$, the joint distribution $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is non-Gaussian. As a result, the optimal estimator $\boldsymbol{\phi}$ cannot be linear.

In the practice of wireless communication, the joint distribution $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is unknown and we must leverage pilot data $\{(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{s}}_1),(\underline{\boldsymbol{x}}_2,\underline{\boldsymbol{s}}_2),\ldots,(\underline{\boldsymbol{x}}_L,\underline{\boldsymbol{s}}_L)\}$ to estimate $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ where L denotes the pilot length. Let

$$\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}} := \frac{1}{L}\sum_{i=1}^{L}\delta_{(\underline{\boldsymbol{x}}_i,\underline{\boldsymbol{s}}_i)} \tag{7}$$

denote the empirical distribution supported on the collected pilot data, which acts as a data-driven estimate of the true but unknown distribution $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$, where $\delta_{(\underline{\boldsymbol{x}}_i,\underline{\boldsymbol{s}}_i)}$ defines the point-mass distribution concentrated at the point $(\underline{\boldsymbol{x}}_i,\underline{\boldsymbol{s}}_i)$. However, the empirical distribution $\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ deviates from the true distribution $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ due to the scarcity of the pilot data. To combat such distributional uncertainty in $\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$, the distributionally robust receive beamforming problem can be formulated as follows:

$$\min_{\boldsymbol{\phi}\in\mathcal{B}}\max_{\mathbb{Q}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}\in\mathcal{U}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}} \operatorname{Tr}\mathbb{E}_{(\underline{\mathbf{x}},\underline{\mathbf{s}})\sim\mathbb{Q}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}}[\boldsymbol{\phi}(\underline{\mathbf{x}})-\underline{\mathbf{s}}][\boldsymbol{\phi}(\underline{\mathbf{x}})-\underline{\mathbf{s}}]^{\mathsf{T}}, \tag{8}$$

where the uncertainty set $\mathcal{U}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ contains a set of distributions that are close to the empirical distribution $\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$; i.e.,

$$\mathcal{U}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}(\theta) := \{\mathbb{Q}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}|\Delta(\mathbb{Q}_{\underline{\mathbf{x}},\underline{\mathbf{s}}},\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}})\le\theta\}, \tag{9}$$

Δ is an appropriate distance (e.g., Wasserstein distance) between $\mathbb{Q}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ and $\hat{\mathbb{P}}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$, and $\theta\ge 0$ indicates the uncertainty level. Although the true data-generating distribution $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is unknown, we assume that $\mathbb{P}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}$ is included in $\mathcal{U}_{\underline{\mathbf{x}},\underline{\mathbf{s}}}(\theta)$ so that (8) can act as a surrogate for, or specifically an upper bound of, (3) in real-world operation.

3. MAIN RESULTS

Outliers in observations $\underline{\mathbf{x}}$ can be caused by several factors such as array abnormality and impulses in channel noises $\underline{\mathbf{v}}$. Since outliers in $\underline{\mathbf{x}}$ are not necessarily from the channel noise $\underline{\mathbf{v}}$, we shall study the non-Gaussianity of $\underline{\mathbf{x}}$ directly.

Suppose that the signal $\underline{\mathbf{s}}$ and the channel noise $\underline{\mathbf{v}}$ are mutually independent. Let $\mathbf{u}\in\mathbb{R}^{2N}$ be defined as

$$\mathbf{u} := \boldsymbol{R}_{\underline{x}}^{-1/2}\cdot\underline{\mathbf{x}}, \tag{10}$$

which is the ***normalized real-space observation***; i.e., the covariance of $\mathbf{u}$ is $\boldsymbol{I}_{2N}$. Therefore, if the received signal $\mathbf{x}$ is complex Gaussian, then $\mathbf{u}$ is a standard multi-variate real Gaussian signal: $\mathbf{u}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_{2N})$. Moreover, $\mathrm{u}_i\sim\mathcal{N}(0,1)$ for every $i\in[2N]$ where u_i is i^{th} component of $\mathbf{u}$, and u_i and u_j are uncorrelated for $i\ne j$; $[2N]:=\{1,2,3,...,2N\}$. In the following discussions, for notational simplicity, we drop the dependence of u_i on $i\in[2N]$.

3.1. Uncertainty Sets

When there exist outliers in u (i.e., in $\mathbf{u}$, $\mathbf{x}$, and $\underline{\mathbf{x}}$), the true fat-tailed distribution of u deviates from the nominal light-tailed Gaussian distribution $\mathcal{N}(0,1)$. Suppose that $F_{\mathrm{u}}(\mu)$ is the cumulative distribution function (CDF) induced by the distribution (i.e., probability measure) $\mathbb{P}_{\mathrm{u}}$ of u. To model outliers in u, two distributional uncertainty sets for $\mathbb{P}_{\mathrm{u}}$ can be used.

1) The ϵ-contamination set [10, Example 4.2]:

$$\mathcal{U}_{\mathrm{u}}(\epsilon) := \left\{\mathbb{P}_{\mathrm{u}}\ \middle|\ \begin{array}{l} F_{\mathrm{u}}(\mu)=\mathbb{P}_{\mathrm{u}}(\mathrm{u}\le\mu) \\ F_{\mathrm{u}}(\mu)=(1-\epsilon)\Phi(\mu)+\epsilon H(\mu) \\ H(\mu)=1-H(-\mu) \\ H(\mu)\text{ is a CDF on }\mathbb{R} \end{array}\right\}, \tag{11}$$

where $\Phi(\mu)$ is the CDF corresponding to $\mathcal{N}(0,1)$; $H(\mu)$ is an arbitrary contamination distribution, which is symmetric about zero. In (11), $F_{\mathrm{u}}(\mu)=(1-\epsilon)\Phi(\mu)+\epsilon H(\mu)$ means that with probability $1-\epsilon$, u is distributed according to the standard Gaussian distribution $\Phi(\mu)$, and with probability ϵ, u is distributed according to a (fat-tailed) contamination distribution $H(\mu)$.

2) The ϵ-normal set [10, Example 4.3]:

$$\mathcal{U}_{\mathrm{u}}(\epsilon) := \left\{\mathbb{P}_{\mathrm{u}}\ \middle|\ \begin{array}{l} F_{\mathrm{u}}(\mu)=\mathbb{P}_{\mathrm{u}}(\mathrm{u}\le\mu) \\ \sup_{\mu\in\mathbb{R}}|F_{\mathrm{u}}(\mu)-\Phi(\mu)|\le\epsilon \\ F_{\mathrm{u}}(\mu)=1-F_{\mathrm{u}}(-\mu) \end{array}\right\}. \tag{12}$$

One can verify that every distribution $\Phi(\mu)$ in (11) satisfies the constraint $\sup_{\mu\in\mathbb{R}} |F_{\text{u}}(\mu) - \Phi(\mu)| \le \epsilon$. Hence, the set in (11) is a subset of that in (12). The difference is that (11) is a structured set, whereas (12) is a general non-structured one.

In practice, the true value of $\boldsymbol{R}_{\underline{x}}$ is unknown and estimated from the collected pilot data. Due to the scarcity of pilot data, the estimated value $\hat{\boldsymbol{R}}_{\underline{x}}$ is also not accurate. Hence, we can construct the uncertainty set for $\hat{\boldsymbol{R}}_{\underline{x}}$ as

$$\mathcal{U}_{R_{\underline{x}}}(\theta_1) := \{\boldsymbol{R}_{\underline{x}} | \Delta_{m,1}(\boldsymbol{R}_{\underline{x}}, \hat{\boldsymbol{R}}_{\underline{x}}) \le \theta_1\}, \tag{13}$$

where $\Delta_{m,1}$ is an appropriate matrix distance between $\boldsymbol{R}_{\underline{x}}$ and $\hat{\boldsymbol{R}}_{\underline{x}}$, and $\theta_1 \ge 0$ quantifies the uncertainty level of $\hat{\boldsymbol{R}}_{\underline{x}}$.

Suppose that the true covariance matrix of the transmitted signal $\underline{\mathbf{s}}$ is $\boldsymbol{R}_{\underline{s}}$. Due to the dynamic time-varying power control operation at the transmitter, the nominally available $\hat{\boldsymbol{R}}_{\underline{s}}$ at the receiver might be different from $\boldsymbol{R}_{\underline{s}}$. Therefore, we construct the uncertainty set for $\hat{\boldsymbol{R}}_{\underline{s}}$ as follows:

$$\mathcal{U}_{R_{\underline{s}}}(\theta_2) := \{\boldsymbol{R}_{\underline{s}} | \Delta_{m,2}(\boldsymbol{R}_{\underline{s}}, \hat{\boldsymbol{R}}_{\underline{s}}) \le \theta_2\}, \tag{14}$$

where $\Delta_{m,2}$ is an appropriate matrix distance between $\boldsymbol{R}_{\underline{s}}$ and $\hat{\boldsymbol{R}}_{\underline{s}}$, and $\theta_2 \ge 0$ quantifies the uncertainty level of $\hat{\boldsymbol{R}}_{\underline{s}}$.

Similarly, the uncertainty set for $\hat{\boldsymbol{R}}_{\underline{xs}}$ can be constructed as

$$\mathcal{U}_{R_{\underline{xs}}}(\theta_3) := \{\boldsymbol{R}_{\underline{xs}} | \Delta_{m,3}(\boldsymbol{R}_{\underline{xs}}, \hat{\boldsymbol{R}}_{\underline{xs}}) \le \theta_3\}, \tag{15}$$

where $\Delta_{m,3}$ is an appropriate matrix distance between $\boldsymbol{R}_{\underline{xs}}$ and $\hat{\boldsymbol{R}}_{\underline{xs}}$, and $\theta_3 \ge 0$ quantifies the uncertainty level of $\hat{\boldsymbol{R}}_{\underline{xs}}$.

Let

$$\boldsymbol{R} := \begin{bmatrix} \boldsymbol{R}_{\underline{x}} & \boldsymbol{R}_{\underline{xs}} \\ \boldsymbol{R}_{\underline{xs}}^{\mathsf{T}} & \boldsymbol{R}_{\underline{s}} \end{bmatrix}. \tag{16}$$

Because $\boldsymbol{R}$ is a covariance matrix, we have $\boldsymbol{R} \succeq \mathbf{0}$.

3.2. Distributionally Robust Outlier-Aware Beamforming

Equipped with the uncertainty assumptions on u, $\hat{\boldsymbol{R}}_{\underline{x}}$, $\hat{\boldsymbol{R}}_{\underline{s}}$, and $\hat{\boldsymbol{R}}_{\underline{xs}}$,[1] the following proposition solves the distributionally robust outlier-aware receive beamforming problem (8), where the type of the channel noise (i.e., $\underline{\mathbf{v}}$) distribution needs not to be explicitly specified.

Proposition 1 *Suppose that $\underline{\mathbf{s}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_{\underline{s}})$, and $\underline{\mathbf{s}}$ and $\underline{\mathbf{v}}$ are uncorrelated. Let* $\mathbf{u}$ *be defined in* (10). *Then the distributionally robust outlier-aware receive beamforming problem* (8) *is particularized into*

$$\min_{\boldsymbol{\phi}\in\mathcal{B}} \max_{\substack{\mathbb{P}_{\text{u}} \in \mathcal{U}_{\text{u}}(\epsilon) \\ \boldsymbol{R}_{\underline{x}} \in \mathcal{U}_{R_{\underline{x}}}(\theta_1) \\ \boldsymbol{R}_{\underline{s}} \in \mathcal{U}_{R_{\underline{s}}}(\theta_2) \\ \boldsymbol{R}_{\underline{xs}} \in \mathcal{U}_{R_{\underline{xs}}}(\theta_3) \\ \boldsymbol{R} \succeq \mathbf{0}}} \operatorname{Tr} \mathbb{E}_{(\mathbf{u},\underline{\mathbf{s}})\sim\mathbb{P}_{\mathbf{u},\underline{\mathbf{s}}}} [\boldsymbol{\phi}(\boldsymbol{R}_{\underline{x}}^{1/2}\mathbf{u}) - \underline{\mathbf{s}}][\boldsymbol{\phi}(\boldsymbol{R}_{\underline{x}}^{1/2}\mathbf{u}) - \underline{\mathbf{s}}]^{\mathsf{T}}. \tag{17}$$

In addition, the distributionally robust outlier-aware estimator of $\underline{\mathbf{s}}$ is given as

$$\hat{\underline{\mathbf{s}}} = \boldsymbol{\phi}(\underline{\mathbf{x}}) = \boldsymbol{R}_{\underline{xs}}^{*\mathsf{T}} \boldsymbol{R}_{\underline{x}}^{*-1/2} \cdot \boldsymbol{\psi}(\boldsymbol{R}_{\underline{x}}^{*-1/2} \underline{\mathbf{x}}) \tag{18}$$

and the covariance of the estimation error is

$$\mathbb{E}[\hat{\underline{\mathbf{s}}} - \underline{\mathbf{s}}][\hat{\underline{\mathbf{s}}} - \underline{\mathbf{s}}]^{\mathsf{T}} = \boldsymbol{R}_{\underline{s}}^* - \boldsymbol{R}_{\underline{xs}}^{*\mathsf{T}} \boldsymbol{R}_{\underline{x}}^{*-1} \boldsymbol{R}_{\underline{xs}}^* \cdot C^{\min}, \tag{19}$$

where

1) *$\boldsymbol{R}_{\underline{x}}^*$, $\boldsymbol{R}_{\underline{s}}^*$, and $\boldsymbol{R}_{\underline{xs}}^*$ solves*

$$\begin{array}{ll} \displaystyle\max_{\boldsymbol{R}_{\underline{x}}, \boldsymbol{R}_{\underline{s}}, \boldsymbol{R}_{\underline{xs}}} & \operatorname{Tr}[\boldsymbol{R}_{\underline{s}} - \boldsymbol{R}_{\underline{xs}}^{\mathsf{T}} \boldsymbol{R}_{\underline{x}}^{-1} \boldsymbol{R}_{\underline{xs}} \cdot C^{\min}] \\ \text{s.t.} & \Delta_{m,1}(\boldsymbol{R}_{\underline{x}}, \hat{\boldsymbol{R}}_{\underline{x}}) \le \theta_1 \\ & \Delta_{m,2}(\boldsymbol{R}_{\underline{s}}, \hat{\boldsymbol{R}}_{\underline{s}}) \le \theta_2 \\ & \Delta_{m,3}(\boldsymbol{R}_{\underline{xs}}, \hat{\boldsymbol{R}}_{\underline{xs}}) \le \theta_3 \\ & \boldsymbol{R}_{\underline{x}} \succ \mathbf{0},\ \boldsymbol{R} \succeq \mathbf{0}. \end{array} \tag{20}$$

2) *If the ϵ-contamination set* (11) *is used, the nonlinear function $\boldsymbol{\psi}(\boldsymbol{\mu})$ is entry-wise identical and for each entry*

$$\psi(\mu) := \begin{cases} -K, & \mu \le -K \\ \mu, & |\mu| \le K \\ K, & \mu \ge K, \end{cases} \tag{21}$$

and the constants K and $C^{\min}$ are implicitly determined by ϵ; see Table 1.

3) *If the ϵ-normal set* (12) *is used, the nonlinear function $\boldsymbol{\psi}(\boldsymbol{\mu})$ is entry-wise identical and for each entry*

$$\psi(\mu) = -\psi(-\mu) := \begin{cases} c\tan(\frac{1}{2}c\mu), & 0 \le \mu \le a \\ \mu, & a \le \mu \le b \\ b, & \mu \ge b, \end{cases} \tag{22}$$

when $\epsilon \le 0.0303$ (i.e., $b \ge a$),

$$\psi(\mu) = -\psi(-\mu) := \begin{cases} c\tan(\frac{1}{2}c\mu), & 0 \le \mu \le a \\ b, & \mu \ge a, \end{cases} \tag{23}$$

when $\epsilon > 0.0303$ (i.e., $b < a$), and the constants a, b, c, and $C^{\min}$ are implicitly determined by ϵ; see Table 2.

Proof: The proof is technically straightforward given the techniques in [11, Subsec. 2.3] and the treatments in [10, Examples 4.2, 4.3]. Therefore, the details of the proof are omitted here; cf. [11, Thm. 8]. □

In Proposition 1, if we let $\theta_1 = \theta_2 = \theta_3 = 0$ and $\epsilon = 0$ (i.e., $\boldsymbol{R}_{\underline{x}}$, $\boldsymbol{R}_{\underline{s}}$, and $\boldsymbol{R}_{\underline{xs}}$ are exactly known and u is strictly Gaussian), we have $\psi(\mu) = \mu$ and (18) becomes the usual linear Wiener beamformer (5). In contrast, if $0 < \epsilon \le 0.5$ (i.e., when large-value outliers exist), the values of the vector $\boldsymbol{R}_{\underline{x}}^{-1/2}\underline{\mathbf{x}}$ are limited (i.e., clipped) to $\pm K$ and $\pm b$; that is, the influence of the large-value outliers in $\underline{\mathbf{x}}$ is limited on the estimator $\boldsymbol{\phi}(\underline{\mathbf{x}})$ of $\underline{\mathbf{s}}$. Therefore, Proposition 1 suggests a nonlinear outlier-thresholded estimation method to recover $\underline{\mathbf{s}}$, i.e., an observation censoring (i.e., clipping) method for $\underline{\mathbf{x}}$. For visual illustrations of the function $\psi(\mu)$, see Fig. 1.

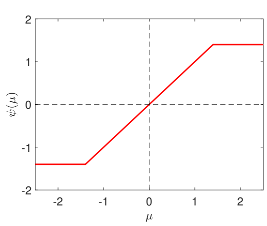
[1] That is, the uncertainty sets $\mathcal{U}_{\text{u}}(\epsilon)$, $\mathcal{U}_{R_{\underline{x}}}(\theta_1)$, $\mathcal{U}_{R_{\underline{s}}}(\theta_2)$, and $\mathcal{U}_{R_{\underline{xs}}}(\theta_3)$, respectively.

Table 1. Constants for ϵ-contamination case.

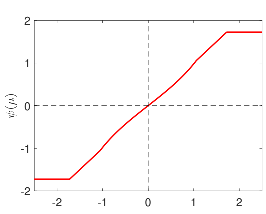
ϵ	K	$1/C^{\min}$
0	∞	1.000
0.001	2.630	1.010
0.002	2.435	1.017
0.005	2.160	1.037
0.01	1.945	1.065
0.02	1.717	1.116
0.05	1.399	1.256
0.10	1.140	1.490
0.15	0.980	1.748
0.20	0.862	2.046
0.25	0.766	2.397

Table 2. Constants for ϵ-normal case.

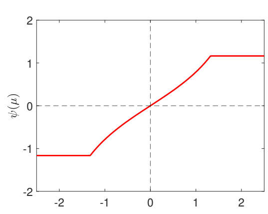
ϵ	a	c	b	$1/C^{\min}$
0	0	1.4142	∞	1.000
0.001	0.6533	1.3658	2.4364	1.019
0.002	0.7534	1.3507	2.2317	1.034
0.005	0.9118	1.3234	1.9483	1.075
0.01	1.0564	1.2953	1.7241	1.136
0.02	1.2288	1.2587	1.4921	1.256
0.03033	1.3496	1.2316	1.3496	1.383
0.05	1.3216	1.1788	1.1637	1.656
0.10	1.3528	1.0240	0.8496	2.613
0.15	1.4335	0.8738	0.6322	4.200
0.20	1.5363	0.7363	0.4674	6.981
0.25	1.6568	0.6108	0.3384	12.24



(a) $\psi(\mu)$ in (21); $\epsilon = 0.05$



(b) $\psi(\mu)$ in (22); $\epsilon = 0.01$



(c) $\psi(\mu)$ in (23); $\epsilon = 0.05$

Fig. 1. Visual illustrations of $\psi(\mu)$. The large values in normalized observations μ are clipped to limit their influences on the estimators, and therefore, the outlier-robustness is achieved; cf. (18). Within the non-clipping interval, $\psi(\mu)$ in (a) is linear, in (b) is hybridly linear and nonlinear (tangent-like), and in (c) is completely nonlinear (tangent-like).

3.3. Solution to (20)

To realize the distributionally robust outlier-aware receiver beamformer (18), the remaining step is to solve maximization problem (20), after particularizing $\Delta_{m,1}$, $\Delta_{m,2}$, and $\Delta_{m,3}$.

Because $\boldsymbol{R}_{\underline{x}}$ and $\boldsymbol{R}_{\underline{s}}$ are positive semidefinite, we can particularize $\Delta_{m,1}(\boldsymbol{R}_{\underline{x}}, \hat{\boldsymbol{R}}_{\underline{x}})$ and $\Delta_{m,2}(\boldsymbol{R}_{\underline{s}}, \hat{\boldsymbol{R}}_{\underline{s}})$ as one of the follows, among many other construction methods.

1) Diagnal Loading [9]:

$$\begin{aligned} \hat{\boldsymbol{R}}_{\underline{x}} - \theta_1 \boldsymbol{I}_{2N} \preceq \boldsymbol{R}_{\underline{x}} \preceq \hat{\boldsymbol{R}}_{\underline{x}} + \theta_1 \boldsymbol{I}_{2N}, \\ \hat{\boldsymbol{R}}_{\underline{s}} - \theta_2 \boldsymbol{I}_{2M} \preceq \boldsymbol{R}_{\underline{s}} \preceq \hat{\boldsymbol{R}}_{\underline{s}} + \theta_2 \boldsymbol{I}_{2M}, \end{aligned} \tag{24}$$

where $\theta_1 \geq 0$ and $\theta_2 \geq 0$ are such that $\hat{\boldsymbol{R}}_{\underline{x}} - \theta_1 \boldsymbol{I}_{2N} \succeq \boldsymbol{0}$ and $\hat{\boldsymbol{R}}_{\underline{s}} - \theta_2 \boldsymbol{I}_{2M} \succeq \boldsymbol{0}$, respectively.

2) Inflating [9]:

$$\begin{aligned} (1-\theta_1)\hat{\boldsymbol{R}}_{\underline{x}} \preceq \boldsymbol{R}_{\underline{x}} \preceq (1+\theta_1)\hat{\boldsymbol{R}}_{\underline{x}}, \\ (1-\theta_2)\hat{\boldsymbol{R}}_{\underline{s}} \preceq \boldsymbol{R}_{\underline{s}} \preceq (1+\theta_2)\hat{\boldsymbol{R}}_{\underline{s}}, \end{aligned} \tag{25}$$

where $\theta_1 \geq 0$ and $\theta_2 \geq 0$ are such that $(1-\theta_1)\hat{\boldsymbol{R}}_{\underline{x}} \succeq \boldsymbol{0}$ and $(1-\theta_2)\hat{\boldsymbol{R}}_{\underline{s}} \succeq \boldsymbol{0}$, respectively.

3) Norm Constraining:

$$\begin{aligned} \|\boldsymbol{R}_{\underline{x}} - \hat{\boldsymbol{R}}_{\underline{x}}\| \leq \theta_1, \\ \|\boldsymbol{R}_{\underline{s}} - \hat{\boldsymbol{R}}_{\underline{s}}\| \leq \theta_2, \end{aligned} \tag{26}$$

where $\|\cdot\|$ denotes any appropriate matrix norm (e.g., F-norm $\|\cdot\|_F$) and $\theta_1, \theta_2 \geq 0$.

Because $\boldsymbol{R}_{\underline{xs}}$ is not square, we can only use norms to particularize $\Delta_{m,3}(\boldsymbol{R}_{\underline{xs}}, \hat{\boldsymbol{R}}_{\underline{xs}})$, i.e.,

$$\|\boldsymbol{R}_{\underline{xs}} - \hat{\boldsymbol{R}}_{\underline{xs}}\| \leq \theta_3. \tag{27}$$

Equipped with the uncertainty assumptions on $\hat{\boldsymbol{R}}_{\underline{x}}$, $\hat{\boldsymbol{R}}_{\underline{s}}$, and $\hat{\boldsymbol{R}}_{\underline{xs}}$, the proposition below solves (20).

Proposition 2 *Problem* (20) *is solved by*

1) *In terms of $\boldsymbol{R}_{\underline{x}}$ and $\boldsymbol{R}_{\underline{s}}$:*

a) *Diagnal Loading:*

$$\begin{aligned} \boldsymbol{R}^*_{\underline{x}} = \hat{\boldsymbol{R}}_{\underline{x}} + \theta_1 \boldsymbol{I}_{2N}, \\ \boldsymbol{R}^*_{\underline{s}} = \hat{\boldsymbol{R}}_{\underline{s}} + \theta_2 \boldsymbol{I}_{2M}, \end{aligned} \tag{28}$$

if the uncertainty set in (24) *is used.*

b) *Inflating:*

$$\begin{aligned} \boldsymbol{R}^*_{\underline{x}} = (1+\theta_1)\hat{\boldsymbol{R}}_{\underline{x}}, \\ \boldsymbol{R}^*_{\underline{s}} = (1+\theta_2)\hat{\boldsymbol{R}}_{\underline{s}}, \end{aligned} \tag{29}$$

if the uncertainty set in (25) *is used.*

2) In terms of $\boldsymbol{R}_{\underline{xs}}$:

$$\begin{aligned} \boldsymbol{R}^*_{\underline{xs}} &= \min_{\boldsymbol{R}_{\underline{xs}}} \quad \mathrm{Tr}[\boldsymbol{R}^\mathsf{T}_{\underline{xs}} \boldsymbol{R}^{*-1}_{\underline{x}} \boldsymbol{R}_{\underline{xs}}] \\ & \quad\; s.t. \quad \|\boldsymbol{R}_{\underline{xs}} - \hat{\boldsymbol{R}}_{\underline{xs}}\| \le \theta_3. \end{aligned} \tag{30}$$

Proof: The objective function in (20) is monotonically increasing in both $\boldsymbol{R}_{\underline{x}}$ and $\boldsymbol{R}_{\underline{s}}$ for every feasible $\boldsymbol{R}_{\underline{xs}}$ because $0 \le C^{\min} \le 1$; cf. Tables 1 and 2. Therefore, the upper bounds of $\boldsymbol{R}_{\underline{x}}$ and $\boldsymbol{R}_{\underline{s}}$ are optimal solutions. Note that all the involved optimization procedures naturally guarantee $\boldsymbol{R} \succeq \boldsymbol{0}$ due to the positive semi-definiteness of the Schur complement of $\boldsymbol{R}$. This completes the proof. □

We do not consider the norm constraints in (26) to obtain $\boldsymbol{R}^*_{\underline{x}}$ and $\boldsymbol{R}^*_{\underline{s}}$ due to technical simplicity because there do not exist closed-form solutions.

Problem (30) in Proposition 2 can be solved by the projected gradient descent method. Note that at the $(t+1)^{\text{th}}$ iteration, the gradient descent step is given by

$$\boldsymbol{R}^{(t+1)}_{\underline{xs}} = \boldsymbol{R}^{*(t)}_{\underline{xs}} - \alpha[(\boldsymbol{R}^{*-1}_{\underline{x}} + \boldsymbol{R}^{*-\mathsf{T}}_{\underline{x}})\boldsymbol{R}^{*(t)}_{\underline{xs}}], \tag{31}$$

where α denotes the step size, and the projection step is given by

$$\begin{aligned} \boldsymbol{R}^{*(t+1)}_{\underline{xs}} &= \min_{\boldsymbol{R}_{\underline{xs}}} \quad \|\boldsymbol{R}_{\underline{xs}} - \boldsymbol{R}^{(t+1)}_{\underline{xs}}\|_F \\ & \quad\; \text{s.t.} \quad \|\boldsymbol{R}_{\underline{xs}} - \hat{\boldsymbol{R}}_{\underline{xs}}\|_F \le \theta_3, \end{aligned} \tag{32}$$

if F-norm is used. By vectorizing the maximization problem in (32), we have

$$\begin{aligned} \min_{\boldsymbol{r}} \quad & \|\boldsymbol{r} - \bar{\boldsymbol{r}}\|_2 \\ \text{s.t.} \quad & \|\boldsymbol{r} - \hat{\boldsymbol{r}}\|_2 \le \theta_3, \end{aligned} \tag{33}$$

where $\boldsymbol{r} := \mathrm{vec}(\boldsymbol{R}_{\underline{xs}})$, $\bar{\boldsymbol{r}} := \mathrm{vec}(\boldsymbol{R}^{(t+1)}_{\underline{xs}})$, and $\hat{\boldsymbol{r}} := \mathrm{vec}(\hat{\boldsymbol{R}}_{\underline{xs}})$. Problem (33) can be analytically solved.

Proposition 3 *If $\|\bar{\boldsymbol{r}} - \hat{\boldsymbol{r}}\|_2 \le \theta_3$, Problem* (33) *is solved by*

$$\boldsymbol{r}^* = \bar{\boldsymbol{r}}; \tag{34}$$

If $\|\bar{\boldsymbol{r}} - \hat{\boldsymbol{r}}\|_2 > \theta_3$, Problem (33) *is solved by*

$$\boldsymbol{r}^* = \gamma\bar{\boldsymbol{r}} + (1-\gamma)\hat{\boldsymbol{r}}, \tag{35}$$

where

$$\gamma := \frac{\theta_3}{\|\bar{\boldsymbol{r}} - \hat{\boldsymbol{r}}\|_2}.$$

Proof: This is obvious due to the geometry of the problem in $\mathbb{R}^{4NM}$ because $\boldsymbol{r}, \bar{\boldsymbol{r}}, \hat{\boldsymbol{r}} \in \mathbb{R}^{4NM}$. □

3.4. Summary of Algorithm

To sum up, the distributionally robust outlier-aware (DROA) receive beamformer is summarized in Algorithm 1.

Algorithm 1 DROA Receive Beamformer

Input: ϵ, θ_1, θ_2, θ_3, T, α

// Obtain $\boldsymbol{R}^*_{\underline{x}}$ and $\boldsymbol{R}^*_{\underline{s}}$; cf. (18) and (19)
  Obtain $\boldsymbol{R}^*_{\underline{x}}$ and $\boldsymbol{R}^*_{\underline{s}}$ using (28) or (29)
// Projected Gradient Descent to solve (30) for $\boldsymbol{R}^*_{\underline{xs}}$
  $t \leftarrow 0$
  $\boldsymbol{R}^{*(t)}_{\underline{xs}} \leftarrow \hat{\boldsymbol{R}}_{\underline{xs}}$
while $t \le T$ or until converges **do**
  // Gradient Descent in (31)
    $\boldsymbol{R}^{(t+1)}_{\underline{xs}} = \boldsymbol{R}^{*(t)}_{\underline{xs}} - \alpha[(\boldsymbol{R}^{*-1}_{\underline{x}} + \boldsymbol{R}^{*-\mathsf{T}}_{\underline{x}})\boldsymbol{R}^{*(t)}_{\underline{xs}}]$
  // Projection
    Solve (32) using Proposition 3
end while
// Determine $\boldsymbol{\psi}(\cdot)$; cf. (18)
  Obtain $\boldsymbol{\psi}(\cdot)$ using (21), (22), or (23); see Tables 1 and 2
// Obtain DROA Receive Beamformer
  Obtain $\underline{\hat{\mathbf{s}}}$ using (18): $\underline{\hat{\mathbf{s}}} = \boldsymbol{R}^{*\mathsf{T}}_{\underline{xs}} \boldsymbol{R}^{*-1/2}_{\underline{x}} \cdot \boldsymbol{\psi}(\boldsymbol{R}^{*-1/2}_{\underline{x}} \underline{\mathbf{x}})$
  Obtain complex-space $\hat{\mathbf{s}}$ using $\underline{\hat{\mathbf{s}}}$; cf. (1) and (2)

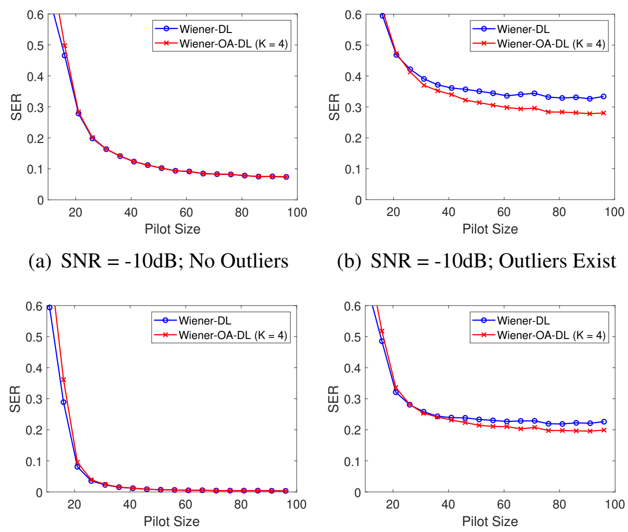
Output: The estimate $\hat{\mathbf{s}}$ of the transmitted signal $\mathbf{s}$

4. EXPERIMENTS

For experimental illustration, we consider a MIMO base-band wireless transmission system where $M = 4$ and $N = 8$.[2] Every data block contains 500 data units for information transmission. In the first scenario, there are no impulse channel noises, and the received signals are not contaminated by outliers; in the second one, however, 10% received signals are contaminated by random outliers due to impulse channel noises. The transmitted signals are quadrature phase-shift keying (QPSK) symbols with unit power, and the symbol error rate (SER) is employed as the performance metric. We assume that the ψ function in (21) is leveraged to suppress outliers, and the diagonal loading scheme with $\theta = 0.05$ in (28) is used to combat the scarcity of pilot data and the uncertainties in the signal model. (We use $\theta = 0.05$ just for a demonstration; one may use other values such as 0.1 or 0.01, which however do not change the main claims in this paper.)

The performances of the outlier-unaware Wiener beamformer (5) and the outlier-robust (i.e., outlier-aware) Wiener beamformer (18) are shown in Fig. 2, against signal-to-noise ratio (SNR) and pilot size L, averaged over 100 Monte–Carlo episodes. The parameter $K = 4$ is empirically tuned since the theoretically optimal value is practically unknown; NB: the true value of ϵ in (11) is practically unknown. As we can see, when outliers are present, 1) SERs significantly rise, and 2) the outlier-aware beamformer in (18) outperforms the outlier-unaware one in (5).

[2] Source codes are available online at GitHub: `https://github.com/Spratm-Asleaf/Beamforming-Outlier`.



(a) SNR = -10dB; No Outliers (b) SNR = -10dB; Outliers Exist

(c) SNR = 0dB; No Outliers (d) SNR = 0dB; Outliers Exist

Fig. 2. Performances of outlier-unaware and outlier-aware (OA) Wiener beamformers against SNR and pilot size L. (DL: Diagonal Loading.)

5. CONCLUSIONS

This paper proposes a distributionally robust outlier-aware receive beamforming method for wireless communications, which is developed from the perspective of uncertainty-aware machine learning: To be specific, the method minimizes the worst-case signal estimation performance to combat the distributional uncertainty in the empirical distribution supported on the pilot data, where the uncertainty is due to the outliers in the received signals, the channel uncertainties, and the scarcity of pilot data. The beamformer essentially employs a nonlinear function to clip the large-valued signals to limit the adverse effects introduced by outliers in the received signals, and the popular diagonal-loading method is a particularization of the proposed beamformer to fight against the channel uncertainties and the scarcity of pilot data. The practical effectiveness of the proposed beamforming method is validated using simulated experiments.

6. REFERENCES

[1] David Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods an results for class a and class b noise models," *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.

[2] T Keith Blankenship, DM Kriztman, and Theodore S Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*. IEEE, 1997, vol. 3, pp. 1942–1946.

[3] Ali A Abdallah and Zaher M Kassas, "Multipath mitigation via synthetic aperture beamforming for indoor and deep urban navigation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8838–8853, 2021.

[4] Victor AN Barroso and José MF Moura, "Maximum likelihood beamforming in the presence of outliers.," in *1991 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1991, pp. 1409–1412.

[5] Bin Li, Yue Rong, Jie Sun, and Kok Lay Teo, "A distributionally robust minimum variance beamformer design," *IEEE Signal Processing Lett.*, vol. 25, no. 1, pp. 105–109, 2017.

[6] Jun Wen, Xiang Zhou, Bin Liao, Chongtao Guo, and Shing-Chow Chan, "Adaptive beamforming in an impulsive noise environment using matrix completion," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 768–771, 2018.

[7] Nadav Yazdi and Koby Todros, "Measure-transformed MVDR beamforming," *IEEE Signal Processing Lett.*, vol. 27, pp. 1959–1963, 2020.

[8] Yongwei Huang, Hao Fu, Sergiy A Vorobyov, and Zhi-Quan Luo, "Robust adaptive beamforming via worst-case SINR maximization with nonconvex uncertainty sets," *IEEE Trans. Signal Processing*, vol. 71, pp. 218–232, 2023.

[9] Shixiong Wang, Wei Dai, and Geoffrey Ye Li, "Distributionally robust beamforming and estimation of wireless signals," *arXiv preprint arXiv:2401.12345*, 2024.

[10] Peter J Huber and Elvezio M Ronchetti, *Robust Statistics*, John Wiley & Sons Hoboken, NJ, USA, 2nd edition, 2009.

[11] Shixiong Wang, *Distributionally Robust State Estimation*, Ph.D. thesis, National University of Singapore, 2022.