Distributionally Robust Adaptive Beamforming

Shixiong Wang, Wei Dai, Geoffrey Ye Li, *Fellow, IEEE*

Abstract—As a fundamental technique in array signal processing, beamforming plays a crucial role in amplifying signals of interest (SoI) while mitigating interference plus noise (IPN). When uncertainties exist in the signal model or the data size of snapshots is limited, the performance of beamformers significantly degrades. In this article, we comprehensively study the conceptual system, theoretical analysis, and algorithmic design for robust beamforming against uncertainties in the assumed snapshot or IPN covariances. Since such robustness is specific to the probabilistic uncertainties of snapshots or IPN signals, it is referred to as distributional robustness. Particularly, four technical approaches for distributionally robust beamforming are proposed, including locally distributionally robust beamforming, globally distributionally robust beamforming, regularized beamforming, and Bayesian-nonparametric beamforming. In addition, we investigate the equivalence among the four technical approaches and suggest a unified distributionally robust beamforming framework. Moreover, we show that the resolution of power spectra estimation using distributionally robust beamforming can be greatly refined by incorporating the characteristics of subspace methods, and hence, the accuracy of IPN covariance reconstruction can be improved, especially when the interferers are close to the SoI. As a result, the robustness of beamformers based on IPN covariance estimation can be further enhanced.

Index Terms—Robust Beamforming, Distributional Robustness, Regularization, Bayesian Nonparametrics

I. INTRODUCTION

Adaptive beamforming, or simply beamforming, has shown numerous successful applications in array signal processing, e.g., wireless communications, radar, and sonar, to enhance signals of interest (SoI) and suppress interference plus noise (IPN). Typical examples include the waveform, power, and direction-of-arrival (DoA) estimation of the SoI, as well as the maximization of the array's output signal-to-interference-plus-noise ratio (SINR) [1]–[4]. To achieve these application purposes, a large body of representative beamformers have been put forward [1], [3], such as the minimum mean-squared error (MMSE) beamformer for waveform estimation, the Bartlett beamformer (i.e., delay-and-sum beamformer, conventional beamformer, or matched filter) for power and DoA estimation, the minimum power distortionless response (MPDR) beamformer (i.e., Capon beamformer) for power and DoA estimation, and the minimum variance distortionless response (MVDR) beamformer for SINR maximization, etc. The difference between the MPDR and MVDR beamformers is whether the snapshot covariance or the IPN covariance is used [5], that is, SoI-contaminated or SoI-free [6]. In the SoI-free case, the snapshot covariance coincides with the IPN covariance, so the MPDR and MVDR beamformers become identical. As a convention and without loss of generality [7,

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom (E-mails: s.wang@u.nus.edu; wei.dai1@imperial.ac.uk; geoffrey.li@imperial.ac.uk).

p. 1705], [8], this article employs SINR as the performance metric for beamformer design, and hence, the MPDR and MVDR beamformers are particularly focused on [9], [10].

A. Uncertainty Issues and Literature Review

The SINR performance of MPDR and MVDR beamformers, however, degrades significantly when uncertainties exist in the assumed steering vector or in the estimated snapshot and IPN covariances [6], [11]–[13], although in such a case the MVDR beamformer tends to outperform the MPDR beamformer [5]. These uncertainties include, for example, array calibration errors, array pointing (i.e., looking-direction) errors, limited data size of snapshots, and non-stationarity of emitting signals or channel noises (e.g., time-varying powers) [2], [12], [13]. To handle these uncertainties, several archetypal treatments have been introduced in the past to improve the robustness of beamformers. Herein, the robustness of a beamformer means the ability to be insensitive to reasonably small uncertainties [6], [9], [11], and consequently, the performance of a robust beamformer in real-world operation can remain as satisfactory as that in the design phase [13]. As per the design motivations, existing treatments can be categorized into three main streams: 1) robustness against uncertainties in steering vectors, 2) robustness against uncertainties in snapshot covariances, and 3) robustness based on IPN covariance reconstruction.

1) Robustness Against Uncertainties in Steering Vectors: Uncertainties in steering vectors come from array calibration errors, pointing errors (i.e., DoA mismatches), etc [6]–[8], [10], [14]. Representatives in this stream include the following:

1a) Linearly constrained minimum variance (LCMV) beamforming [15] [13, Section 2.5], [16, Section 6];

1b) Covariance matrix tapering to widen nulls and beams, which leverages the phase dithering effect [12], [17, Section 2]. Some technical equivalences between this approach and Method 1a) are noted in [18];

1c) Bayesian beamforming, which assumes a probability distribution of the DoA, and then obtains the expected beamformer under this DoA distribution [19], [20];

1d) Beampattern control, which, conceptually similar to Method 1a), requires the array power response to comply with some prerequisites [16], [21];

1e) Steering vector estimation via eigenspace projection [22], [23] or output power maximization [14], [24], [25]. The eigenspace projection indirectly increases the array output power in the MPDR beamforming scheme. Also, the covariance-fitting methods in [7, Eqs. (8), (9)] and [26, Eqs. (10), (13)] amount to output power maximization;

1f) Worst-case array unity response [6]–[8], [10], [17], [27]. Regarding objective functions, Method 1e) is technically equivalent to this approach under the MPDR beamforming

scheme; see [9, Eq. (17)], [14, Eq. (23)], [25, Eq. (12)], [7, Appendices A, C];

1g) Probabilistic worst-case array unity response, which can be technically cast into Method 1f) [28];

1h) Imposing quadratic constraints on beamformers [11, p. 1366], [29, p. 742]. Some beampattern control problems are technically equivalent to this approach [16, Eq. (22)].

Methods 1a), 1b), and 1c) are ad hoc to DoA mismatches, while Methods 1d), 1e), 1f), 1g), and 1h) can apply to arbitrary types of uncertainties in the steering vector. The philosophy of robustness behind Methods 1a), 1b), 1d), 1f), and 1g) is to identify a plausible uncertainty set for the actual but unknown steering vector, and then guarantee that the array power responses for all elements in this set are controlled, for example, greater than unity. In the case of DoA mismatch, the uncertainty set is induced by the uncertainty region of DoA under the array geometry. In contrast, the philosophy of robustness behind Method 1c) is to consider an uncertainty region of DoA through employing a distribution of DoA. Moreover, the philosophy of robustness behind Method 1e) is that the actual steering vector should lie in the subspace of signal-plus-interference or should increase the array output power. In addition, the philosophy of robustness behind Method 1h) is to limit the array's sensitivity to various errors because the sensitivity can be measured by the 2-norm of beamformers.

All methods, except Method 1e), can be integrated or adapted into both the MPDR and MVDR beamforming to enhance their robustness. Method 1e) cannot be considered in the MVDR beamforming scheme because in this case, the subspace of signal-plus-interference and the array output power are not applicable.

2) Robustness Against Uncertainties in Snapshot Covariances: Uncertainties in snapshot covariances can be caused by the limited data size of snapshots, the non-stationarity of signal characteristics, etc [10], [12], [17], [30]–[32]. When the SoI is absent in snapshots, the snapshot covariance coincides with the IPN covariance. Representatives in this stream include the following:

2a) Advanced techniques to estimate the covariance matrix using snapshots, for example, the M-estimator [33], [34], the spiked covariance estimator [32], the eigenvalue thresholding method [35], the diagonal loading method [30], [36]–[38], the prior-knowledge embedding method [31], etc. The M-estimator can suppress outliers in snapshots, while the remaining can combat snapshot scarcity.

2b) Worst-case optimization [17, Eq. (46)], [10, Eq. (10)], [9, Eq. (9)]. This approach can be shown to technically amount to the diagonal loading method [17].

The philosophy of robustness behind Method 2a) is to obtain a better covariance estimate, which is closer to the ground truth than the usual sample covariance matrix. By feeding this better covariance to beamforming, the performance is expected to improve. On the contrary, the philosophy of robustness behind Method 2b) is to optimize the worst-case performance, e.g., to maximize the worst-case SINR [9]. In [10], [17], this worst-case SINR is determined by minimizing SINR over the covariance estimation error, which is characterized by an F-norm ball.

Depending on whether the SoI is present in snapshots or not, these methods correspond to the robust MPDR and MVDR beamformers, respectively.

3) Robustness Based on IPN Covariance Reconstruction: Motivated by the fact that the MVDR beamformer tends to be more robust than the MPDR beamformer [5], another philosophy to achieve robustness is to estimate the IPN covariance matrix from snapshots before beamforming. This approach applies to the case where the SoI is present in snapshots. Representatives in this stream include the following:

3a) Strategies via spectra estimation, for example, the Capon spectra [39] and the maximum-entropy spectra [40]. Improvements in terms of uncertainty-awareness and computation reduction are reported in [41] and [42], [43], respectively;

3b) Strategies via SoI cancellation, e.g., [44, Eq. (26)] and [45, Eq. (27)], where the snapshot covariance is first estimated and then the IPN covariance is recovered through subtracting the SoI component from the snapshot covariance.

B. Problem Statements

Although various challenges in robust adaptive beamforming have been attacked, the following issues remain unsolved.

I1) In the literature, although the notion of "robust beamforming" is widely used, the quantification and formalization have never been studied. The only effort in this direction is found in, e.g., [11], [29], where the authors propose to use the white-noise gain [11] or the (normalized) squared magnitude [29] as a quantitative measure of the robustness of a beamformer. However, this robustness formalism does not apply to the uncertainties in the snapshot and IPN covariances. On the other hand, although numerous minimax-optimization-based approaches claim the robustness of their resultant beamformers [17, Eq. (46)], [10, Eq. (10)], the rigorous relation between the beamformer's robustness and minimax optimization remains unclear. Therefore, a comprehensive study of the conceptual system, theoretical analysis, and algorithmic design for robust beamforming, *specifically against uncertainties in snapshot and IPN covariances*, has to be conducted.

I2) Regarding robustness against uncertainties in snapshot covariances, the relations between Methods 2a) and 2b) are unclear. To clarify further, in the existing literature, researchers tacitly employ the eigenvalue thresholding method, the diagonal loading method, or the prior-knowledge embedding method, and then focus on tuning the involved parameters using different empirical or algorithmic criteria [30], [31], [36]–[38]. However, the rationale behind these choices has not been systematically elaborated. A specific question is as follows: In addition to diagonal loading [17], can eigenvalue thresholding [35] and prior-knowledge embedding [31] also be shown to optimize the worst-case performance? This issue remains valid for both the MPDR beamformer that relies on the snapshot covariance and the MVDR beamformer that relies on the IPN covariance.

I3) In the robustness framework based on IPN covariance reconstruction, i.e., Methods 3a) and 3b), the estimation error of IPN is also unavoidable. Hence, it is natural to ask: Can the robustness of this framework be further improved using the robustness strategies in Methods 2a) and 2b)? This question has not been answered in the literature.

I4) In Method 3a) for IPN covariance reconstruction, the resolution of spectra estimation should be sufficiently high to suppress interferers that locate close to the SoI. This concern has also been raised in [41, p. 1647], [40, Fig. 1]. However, the resolution of the Capon spectra and the maximum-entropy spectra are still limited [46, Fig. 3], [47], and therefore, higher-resolution spectra estimation methods are expected to improve the estimation accuracy of the IPN covariance, especially when interferers are close to the SoI. The challenge is that modern high-resolution spectra estimation methods, such as multiple signal classification (MUSIC) and other subspace strategies, do not directly provide spectral estimates. Instead, they generate pseudo-spectra to identify the DoAs of interferers. On the other hand, to ensure high accuracy of IPN covariance reconstruction, the power spectra estimation should be unbiased, that is, no gaps between identified DoAs and true DoAs of interferers. However, the maximum-entropy spectra tend to have large biases [46, Fig. 4]. As for Method 3b), it innately assumes that interferers are well-separated from the SoI, which limits its applicability when interferers are close to the SoI.

C. Contributions

To address the four issues aforementioned, this article makes the following contributions.

C1) We propose a quantitative definition of robustness *against uncertainties in snapshot and IPN covariances*, and prove that many existing beamformers in Methods 2a) and 2b) are indeed robust; for example, diagonal loading, eigenvalue thresholding, prior-knowledge embedding can be shown to amount to worst-case optimization; see Definitions 1 and 2, Theorems 1, 2, and 3, Examples 1, 2, 3, 4, and 6, Corollaries 1 and 2, and Insights 1 and 2. To emphasize the probabilistic nature of snapshots and IPN signals, and differentiate from the robustness against uncertainties in steering vectors, we call the introduced concept as "*distributional robustness*". Subsequently, four technical approaches for distributionally robust beamforming are proposed, i.e., locally distributionally robust beamforming, globally distributionally robust beamforming, regularized beamforming, and Bayesian-nonparametric beamforming; see Models (20), (31), (41), and (45). The equivalence among these approaches is investigated and a unified distributionally robust beamforming framework is suggested; see Insight 3.

C2) We show that by incorporating the characteristics of the MUSIC method into the proposed distributionally robust (DR) beamforming framework, the resolution of spectra estimation can be greatly refined. As a result, the SINR performance of DR beamformers based on IPN covariance reconstruction can be largely improved when interferers are close to the SoI; see Insight 4, Method 1, and Algorithm 1.

Contribution C1) solves Issues I1), I2), and I3), while Contribution C2) addresses Issue I4).

D. Notations

Uppercase symbols (e.g., $\boldsymbol{X}$) denote matrices while lowercase ones are reserved for vectors (e.g., $\boldsymbol{x}$). We use upright and italic fonts for random and deterministic quantities, respectively; e.g., $\mathbf{X}$ and $\mathbf{x}$, and $\boldsymbol{X}$ and $\boldsymbol{x}$. Let $\mathbb{C}^d$ denote the d-dimensional space of complex numbers. The running index set $[K]$ induced by integer K is defined as $[K] := \{1, 2, 3, \ldots, K\}$. Let $\|\boldsymbol{X}\|$, $\mathrm{Tr}\,\boldsymbol{X}$, $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^{\mathsf{T}}$, and $\boldsymbol{X}^{\mathsf{H}}$ denote a norm, the trace, the inverse (if exists), the transpose, and the conjugate transpose of matrix $\boldsymbol{X}$; the definition of a matrix norm will be specified in contexts. For two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite. The d-dimensional identity matrix is written as $\boldsymbol{I}_d$ and a zero matrix/vector with compatible dimensions is as $\boldsymbol{0}$. Let $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}')$ denote the complex normal distribution with mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}$, and pseudo-covariance $\boldsymbol{\Sigma}'$; if $\boldsymbol{\Sigma}'$ is not specified, we admit $\boldsymbol{\Sigma}' = \boldsymbol{0}$. Let $\mathbb{E}_{\mathbb{P}}[\cdot]$ denote the expectation operator under distribution $\mathbb{P}$.

II. SYSTEM MODEL, EXISTING WORKS, AND RESEARCH QUESTIONS

A. System Model

This article focuses on base-band and narrow-band signal processing. Suppose K signals impinge on an N-element antenna array. Let $\mathbf{x} \in \mathbb{C}^N$ denote the array input (i.e., snapshot), $\mathrm{s}_k \in \mathbb{C}$ the kth incident signal for $k \in [K]$, and θ_k the DoA of s_k. The signal model for a single snapshot is

$$\mathbf{x} = \sum_{k=1}^{K} \boldsymbol{a}_0(\theta_k)\mathrm{s}_k + \mathbf{v}, \tag{1}$$

where $\boldsymbol{a}_0(\theta) \in \mathbb{C}^N$ denotes the array steering vector in direction θ and $\mathbf{v} \in \mathbb{C}^N$ the channel noise. In usual array signal processing literature, it is assumed that $\mathrm{s}_k \sim \mathcal{CN}(0, \sigma_k^2)$ and $\mathbf{v} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_N)$, in which σ_k^2 is the signal power and σ_n^2 is the noise power. Without loss of generality, we suppose that s_1 is the signal of interest and s_k for $k = 2, 3, \ldots, K$ are interference signals. Let $\boldsymbol{w} \in \mathbb{C}^N$ be a beamformer. The corresponding array output is

$$\mathrm{y} = \boldsymbol{w}^{\mathsf{H}}\mathbf{x} = \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0(\theta_1)\mathrm{s}_1 + \boldsymbol{w}^{\mathsf{H}}\sum_{k=2}^{K}\boldsymbol{a}_0(\theta_k)\mathrm{s}_k + \boldsymbol{w}^{\mathsf{H}}\mathbf{v}, \tag{2}$$

and the array output SINR is

$$h(\boldsymbol{w}, \boldsymbol{a}_0(\theta_1), \boldsymbol{R}_{i+n}) := \frac{\sigma_1^2 \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0(\theta_1)\boldsymbol{a}_0(\theta_1)^{\mathsf{H}}\boldsymbol{w}}{\boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}_{i+n}\boldsymbol{w}}, \tag{3}$$

where $\boldsymbol{R}_{i+n} := \sum_{k=2}^{K}\sigma_k^2\boldsymbol{a}_0(\theta_k)\boldsymbol{a}_0(\theta_k)^{\mathsf{H}} + \sigma_n^2\boldsymbol{I}_N$ denotes the covariance of the interference signals plus noise. Depending on specific applications, the typical roles of beamforming include the following: to adjust array output SINR (3), to estimate SoI

waveform s_1 using y [4], [48], to estimate SoI power σ_1^2 using $\mathbb{E}yy^{\mathsf{H}}$ [7], [26], and to estimate DoAs θ_k using the directions corresponding to the largest output powers $\mathbb{E}yy^{\mathsf{H}}$ [47, p. 1019].

The MVDR beamformer in direction θ solves the following beamforming problem:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}_{i+n}\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0(\theta) = 1. \end{aligned} \tag{4}$$

The MVDR beamformer is optimal in many senses: to achieve maximum output SINR (3) and to attain maximum likelihood estimate of SoI s_1 given x when $\theta := \theta_1$ [47, Eq. (20)]. In practice, however, the IPN covariance $\boldsymbol{R}_{i+n}$ is unknown, and alternatively, the following MPDR (i.e., Capon) beamforming problem that minimizes the array output power $\mathbb{E}yy^{\mathsf{H}}$ in direction θ is solved:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}_x\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0(\theta) = 1, \end{aligned} \tag{5}$$

where $\boldsymbol{R}_x := \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_{\mathbf{x}}}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]$ denotes the covariance of received signal x in direction θ and $\mathbb{P}_{\mathbf{x}}$ its underlying true distribution; note that $\boldsymbol{R}_x$ can be estimated using collected snapshots. In what follows, we use $\boldsymbol{a}_0$ as a shorthand for $\boldsymbol{a}_0(\theta)$, if no ambiguity is caused. Among existing beamformers, when the snapshot covariance $\boldsymbol{R}_x$ and the array steering vector $\boldsymbol{a}_0$ are (almost) exactly known, the MPDR beamformer is most popular for its excellent performance and real-world operationality. To be specific, the MPDR beamformer is optimal in the sense of MPDR waveform estimation [8], [4, Eq. (4)], MPDR power estimation [7, Eq. (7)], high-resolution DoA estimation [47], [49], and SINR maximization [9, p. 1540], [13, Eq. (10)]. In addition, the snapshot covariance can be estimated using the array input data, while the IPN covariance in the MVDP beamforming is difficult to obtain for the SoI-contaminated case. However, if the SoI s_1 is absent in the snapshot x, the snapshot covariance $\boldsymbol{R}_x$ matches the IPN covariance $\boldsymbol{R}_{i+n}$, and the MPDR beamformer becomes the MVDR beamformer.

Hereafter, according to whether the MPDR or MVDR beamforming scheme is used, we let $\boldsymbol{R}_0$ denote either the snapshot covariance $\boldsymbol{R}_x$ or the IPN covariance $\boldsymbol{R}_{i+n}$. Let $\hat{\boldsymbol{R}}$, $\hat{\boldsymbol{R}}_x$, $\hat{\boldsymbol{R}}_{i+n}$, and $\hat{\boldsymbol{a}}$ denote the estimates of $\boldsymbol{R}_0$, $\boldsymbol{R}_x$, $\boldsymbol{R}_{i+n}$, and $\boldsymbol{a}_0$, respectively.

B. Existing Works

To address the uncertainty in $\hat{\boldsymbol{R}}$ compared to its true value $\boldsymbol{R}_0$, typical treatments include 1) eigenvalue thresholding, 2) diagonal loading, and 3) prior-knowledge embedding, 4) worst-case optimization, and 5) IPN covariance reconstruction.

Eigenvalue Thresholding: Let $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ be eigenvalues of $\hat{\boldsymbol{R}}$ in descending order and $\hat{\boldsymbol{U}}$ contains eigenvectors. Define $\hat{\boldsymbol{R}}_{\text{thr}} := \hat{\boldsymbol{U}}\boldsymbol{\Lambda}_{\text{thr}}\hat{\boldsymbol{U}}^{\mathsf{H}}$ and

$$\boldsymbol{\Lambda}_{\text{thr}} := \begin{bmatrix} \lambda_1 & & & \\ & \max\{\mu\lambda_1, \lambda_2\} & & \\ & & \ddots & \\ & & & \max\{\mu\lambda_1, \lambda_N\} \end{bmatrix} \tag{6}$$

where $0 \le \mu \le 1$; NB: when $\mu = 0$, $\hat{\boldsymbol{R}}_{\text{thr}}$ reduces to $\hat{\boldsymbol{R}}$. The eigenvalue thresholding method to combat the uncertainty in $\hat{\boldsymbol{R}}$ is to use $\hat{\boldsymbol{R}}_{\text{thr}}$ in beamforming [35, Eq. (36)], [8, Eq. (12)].

Diagonal Loading: The diagonal loading method to combat the uncertainty in $\hat{\boldsymbol{R}}$ is to use $\hat{\boldsymbol{R}}+\epsilon\boldsymbol{I}_N$ in beamforming, where $\epsilon \ge 0$ is a scalar [30], [36]–[38]. Different tuning principles for ϵ have been discussed, including main beam correction and side lobe reduction [12], [30], asymptotic properties [36], empirical mean-squared error (MSE) minimization [31], SoI power estimation [37], etc. An interesting observation is that some robust methods against the uncertainty in the assumed steering vector $\hat{\boldsymbol{a}}$ can be shown as the diagonal loading method, where ϵ is determined by the uncertainty degree of $\hat{\boldsymbol{a}}$ [6], [7] or the sensitivity measure of the array [11], [29]. In practice, however, the best tuning method for ϵ is still trial-and-error because the best value under one criterion challenges the optimality (or even satisfaction) under the other criterion.

Prior-Knowledge Embedding: The prior-knowledge embedding method to combat the uncertainty in $\hat{\boldsymbol{R}}$ is to use $\alpha\hat{\boldsymbol{R}}+\beta\bar{\boldsymbol{R}}$ in beamforming, for some weight coefficients $\alpha, \beta \ge 0$ and prior knowledge $\bar{\boldsymbol{R}}$ of $\boldsymbol{R}_0$ [31]. It is believed that $\alpha\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}}$ can provide a better estimate than $\hat{\boldsymbol{R}}$ and $\bar{\boldsymbol{R}}$, in the sense of smaller MSE. The coefficients α and β can be tuned using empirical MSE minimization [31].

Worst-Case Optimization: Robust beamforming in the sense of minimax optimization is formulated as [6], [8], [10], [17]

$$\begin{aligned} \min_{\boldsymbol{w}} \max_{\boldsymbol{R} \in \mathcal{U}_R} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w} \\ \text{s.t.} \quad & \min_{\boldsymbol{a} \in \mathcal{U}_a} \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\boldsymbol{w} \ge 1, \end{aligned} \tag{7}$$

where $\mathcal{U}_R$ and $\mathcal{U}_a$ are the uncertainty sets for $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{a}}$, respectively; $\hat{\boldsymbol{R}}, \boldsymbol{R}_0 \in \mathcal{U}_R$ and $\hat{\boldsymbol{a}}, \boldsymbol{a}_0 \in \mathcal{U}_a$. The philosophy behind (7) is to *optimize the worst-case performance and guarantee the worst-case feasibility*. When $\mathcal{U}_R := \{\boldsymbol{R} : \|\boldsymbol{R}-\hat{\boldsymbol{R}}\|_F \le \epsilon_1\}$ and $\mathcal{U}_a := \{\boldsymbol{a} : \|\boldsymbol{a} - \hat{\boldsymbol{a}}\|_2 \le \epsilon_2\}$ for some $\epsilon_1, \epsilon_2 \ge 0$ where $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the matrix Frobenius norm and the vector 2-norm, respectively, Problem (7) can be equivalently transformed into [6, Eq. (29)], [10, Eq. (13)]

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}(\hat{\boldsymbol{R}} + \epsilon_1\boldsymbol{I}_N)\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{a}} \ge \epsilon_2\|\boldsymbol{w}\|_2 + 1, \end{aligned} \tag{8}$$

which is tantamount to the diagonal loading method in terms of the objective function. Problem (8) is equivalent, in the sense of the same optimal cost, to

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}(\hat{\boldsymbol{R}} + \epsilon_1\boldsymbol{I}_N)\boldsymbol{w} \\ \text{s.t.} \quad & |\boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{a}}| \ge \epsilon_2\|\boldsymbol{w}\|_2 + 1, \end{aligned} \tag{9}$$

because if $\boldsymbol{w}^*$ is an optimal solution to (9), so is $\boldsymbol{w}^*e^{j\varphi}$ for any angle φ; j denotes the imaginary unit; note that $\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}$ is rotation-invariant. For other proposals of $\mathcal{U}_a$, see, e.g., [50, p. 871], [51, p. 2408], [8, Eq. (13)], [14, Eq. (25)], [39, Eq. (10)], [10, p. 221]. Note that the robustification in terms of $\boldsymbol{R}$ and $\boldsymbol{a}$ can be independently conducted; cf. (7). Under the MVDR beamforming scheme, which occurs when the SoI is not included in the snapshots or when the IPN covariance $\boldsymbol{R}_{i+n}$ is estimated using Methods 3a) or 3b) [39], [40], [44], another interpretation of (7) is to maximize the worst-case SINR [9], [10]; i.e., $\max_{\boldsymbol{w}} \min_{\boldsymbol{a},\boldsymbol{R}} h(\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{R})$; cf. (3).

IPN Covariance Reconstruction: The motivation behind this method is that the MVDR beamformer tends to be more robust than the MPDR beamformer [5]. Hence, in the SoI-contaminated case, a natural way is to estimate the IPN covariance using snapshots [39], [40], [44], [45]. In the typical treatment, the key step is to estimate the power spectra $P(\theta)$ and then reconstruct the IPN covariance as follows

$$\hat{\boldsymbol{R}}_{i+n} = \int_{\bar{\Theta}} P(\theta)\boldsymbol{a}_0(\theta)\boldsymbol{a}_0^{\mathsf{H}}(\theta)\mathrm{d}\theta, \qquad (10)$$

where $\bar{\Theta}$ is the angular sector that excludes the uncertainty region of the SoI's DoA [39], [40]. In real-world operation, $P(\theta)$ is estimated using the Capon spectra [39] or the maximum-entropy spectra [40], and $\boldsymbol{a}_0(\theta)$ is estimated using $\hat{\boldsymbol{a}}(\theta)$.

C. Specific Research Questions

From the review of existing works in Subsection II-B, we can see that the following questions have not been answered.

Q1) According to [11, p. 1365], [6, p. 313], [29, p. 742], and our intuition, the well accepted notion of "robust beamforming" means that the beamformer is insensitive to possible perturbations in signal characteristics. However, how does the worst-case optimization method in (7) reflect this notion in terms of the uncertainty in $\hat{\boldsymbol{R}}$?

Q2) Can we show that the existing eigenvalue thresholding method and the prior-knowledge embedding method are also tantamount to the worst-case optimization method (7), as the diagonal loading method is?

Q3) In IPN covariance reconstruction, the resolution and bias of spectra estimation is crucial, especially for suppressing the closely located interferers. Therefore, how can we incorporate the higher-resolution and lower-bias spectra estimation methods, such as MUSIC, into IPN covariance reconstruction?

The above three questions are representative particularizations of identified Issues I1)-I4) in Subsection I-B. Note that Questions Q1) and Q2) apply to both MPDR and MVDR beamforming.

III. DISTRIBUTIONALLY ROBUST ADAPTIVE BEAMFORMING

This section studies the formalized theory of distributionally robust (adaptive) beamforming, with a focus on combating the uncertainties in the estimated snapshot or IPN covariances. In particular, the concept of "*distributional robustness*" is quantitatively defined and the countermeasure methods are proposed. The concept of distributional robustness is noted because we are working with the probabilistic uncertainties in snapshots or IPN signals. This stochasticity feature fundamentally differs from the fixed (albeit unknown) uncertainty in the assumed steering vector $\hat{\boldsymbol{a}}$.

In view of potential errors in the assumed steering vector $\hat{\boldsymbol{a}}$, we consider the feasible set of beamformers $\boldsymbol{w}$ as

$$\mathcal{W} := \left\{ \boldsymbol{w} : \min_{\boldsymbol{a} \in \mathcal{U}_a} \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\boldsymbol{w} \geq 1 \right\}, \qquad (11)$$

for a given uncertainty set $\mathcal{U}_a$ of the steering vector. However, to focus on the uncertainty in $\hat{\boldsymbol{R}}$, the uncertainty in $\hat{\boldsymbol{a}}$ is minimally examined in this article.

A. Issue of Distributional Uncertainty

We revisit the distributional form of beamforming

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}\mathbb{E}_{\mathbf{s}\sim\mathbb{P}_0}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0 = 1, \end{aligned} \qquad (12)$$

where $\mathbf{s}$ denotes the snapshot $\mathbf{x}$ or the IPN signal under the MPDR and MVDR beamforming, respectively, and $\mathbb{P}_0$ the underlying true distribution of $\mathbf{s}$. Since $\mathbb{P}_0$ is unavailable in practice, the empirical distribution

$$\hat{\mathbb{P}} := \frac{1}{L}\sum_{l=1}^{L} \delta_{\boldsymbol{s}_l} \qquad (13)$$

constructed using L collected samples $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_L\}$ can serve as an estimate of $\mathbb{P}_0$ where $\delta_{\boldsymbol{s}}$ denotes the point-mass distribution centered at $\boldsymbol{s}$. As a result, the sample-average approximation (SAA) of (12) can be written as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{\mathsf{H}}\mathbb{E}_{\mathbf{s}\sim\hat{\mathbb{P}}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}_0 = 1. \end{aligned} \qquad (14)$$

When $\mathcal{U}_a$ contains only $\boldsymbol{a}_0$,

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\mathbb{E}_{\mathbf{s}\sim\mathbb{P}_0}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]\boldsymbol{w} \qquad (15)$$

reduces to (12) and

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\mathbb{E}_{\mathbf{s}\sim\hat{\mathbb{P}}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]\boldsymbol{w} \qquad (16)$$

reduces to (14). Since $\hat{\mathbb{P}}$ is distributionally uncertain compared to $\mathbb{P}_0$, directly employing (16) as a surrogate of (15) is questionable due to the issue of "overfitting on data".

In beamforming, as only the second moments of $\hat{\mathbb{P}}$ and $\mathbb{P}_0$ are involved, for presentation simplicity, we directly work on $\hat{\boldsymbol{R}} := \mathbb{E}_{\mathbf{s}\sim\hat{\mathbb{P}}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]$ and $\boldsymbol{R}_0 := \mathbb{E}_{\mathbf{s}\sim\mathbb{P}_0}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]$ whenever possible. Only when technically necessary, we investigate $\hat{\mathbb{P}}$ and $\mathbb{P}_0$. Another benefit of this treatment is that, under the MVDR beamforming scheme for SoI-contaminated cases, the IPN signal $\mathbf{s}$ cannot be directly observed but its covariance can still be estimated [39], [45].

B. Definition of Distributional Robustness

Motivated by the philosophical notion of robustness, we study the formal definitions of distributional robustness. We begin with the concept of local distributional robustness.

Definition 1 (Locally Distributionally Robust Beamformer): A beamformer $\boldsymbol{w}^*$ is called (ϵ, k)-locally-robust on the uncertainty set $\mathcal{B}_\epsilon(\boldsymbol{R}_0) := \{\boldsymbol{R} : d(\boldsymbol{R}, \boldsymbol{R}_0) \leq \epsilon\}$ if

$$\boldsymbol{w}^{*\mathsf{H}}\boldsymbol{R}\boldsymbol{w}^* - \boldsymbol{w}_0^{\mathsf{H}}\boldsymbol{R}_0\boldsymbol{w}_0 \leq k, \qquad \forall \boldsymbol{R} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0), \qquad (17)$$

where $\boldsymbol{w}_0 \in \min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}_0\boldsymbol{w}$ is an optimal beamformer associated with $\boldsymbol{R}_0$; d is a matrix similarity measure. The smallest value k^* of $k \geq 0$ satisfying (17) is called the *local robustness measure* of $\boldsymbol{w}^*$ at ϵ. □

Definition 1 means that when the real-world operating covariance $\boldsymbol{R}$ deviates from the underlying true $\boldsymbol{R}_0$, the performance degradation $\boldsymbol{w}^{*\mathsf{H}}\boldsymbol{R}\boldsymbol{w}^* - \boldsymbol{w}_0^{\mathsf{H}}\boldsymbol{R}_0\boldsymbol{w}_0$ at the robust beamformer $\boldsymbol{w}^*$ is upper bounded by k and k^*. This formalism straightforwardly reflects the notion of robustness, that is, the

insensitivity in terms of perturbations in the covariance matrix. Note that the smaller the value of k and k^*, the smaller the performance degradation under uncertainties, and therefore, the more robust the beamformer $\boldsymbol{w}^*$ is.

Definition 1 is natural if the deviation level ϵ is known, which, however, is not always the case in practice. Hence, the concept of global distributional robustness can be motivated.

Definition 2 (Globally Distributionally Robust Beamformer): A beamformer $\boldsymbol{w}^*$ is called (τ, k)-globally-robust on the whole space $\mathbb{C}^{N \times N}$ if

$$\boldsymbol{w}^{*\mathsf{H}} \boldsymbol{R} \boldsymbol{w}^* - \tau \le k \cdot d(\boldsymbol{R}, \boldsymbol{R}_0), \qquad \forall \boldsymbol{R} \in \mathbb{C}^{N \times N}, \quad (18)$$

where τ is a prescribed cost threshold satisfying $\boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0 \le \tau < \infty$. The smallest value k^* of $k \ge 0$ satisfying (18) is called the *global robustness measure* of $\boldsymbol{w}^*$ at τ. □

Definition 2 means that when the real-world operating covariance $\boldsymbol{R}$ deviates from the underlying true $\boldsymbol{R}_0$, the performance degradation $\boldsymbol{w}^{*\mathsf{H}} \boldsymbol{R} \boldsymbol{w}^* - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0$ at the robust beamformer $\boldsymbol{w}^*$ is upper bounded by $\tau - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0 + k \cdot d(\boldsymbol{R}, \boldsymbol{R}_0)$. For modeling flexibility, we do not restrict $\tau \equiv \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0$.

Hereafter in this article, for presentation brevity, we shorthand "distributional robustness" simply as "robustness", if no ambiguity is caused. In addition, we do not explicitly mention whether the referred robustness is local or global; this can be straightforwardly inferred from context.

C. Locally Distributionally Robust Beamforming

In line with Definition 1 and its motivation, we propose the following robust beamforming formulation

$$
\begin{aligned}
(\boldsymbol{w}^*, k^*) = \;& \underset{\boldsymbol{w} \in \mathcal{W}, k}{\operatorname{argmin}} \quad k \\
& \text{s.t.} \qquad \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0 \le k, \quad \forall \boldsymbol{R} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0), \\
& \qquad\quad k \ge 0,
\end{aligned}
$$
$$(19)$$

which finds the locally robust beamformer $\boldsymbol{w}^*$ and the local robustness measure k^*. In practice where $\boldsymbol{R}_0$ is inaccessible, we can solve (19) resorting to

$$
\begin{aligned}
\min_{\boldsymbol{w} \in \mathcal{W}, k} \quad & k \\
\text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - \hat{\boldsymbol{w}}^{\mathsf{H}} \hat{\boldsymbol{R}} \hat{\boldsymbol{w}} \le k, \quad \forall \boldsymbol{R} \in \mathcal{B}_\delta(\hat{\boldsymbol{R}}), \\
& k \ge 0,
\end{aligned}
$$
$$(20)$$

where $\hat{\boldsymbol{R}}$ is an estimate of $\boldsymbol{R}_0$ and

$$\hat{\boldsymbol{w}} \in \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{R}} \boldsymbol{w} \qquad (21)$$

is an optimal beamformer associated with $\hat{\boldsymbol{R}}$. The intuition is that if $\boldsymbol{R}_0 \in \mathcal{B}_\delta(\hat{\boldsymbol{R}})$ for some $\delta \ge 0$, then there exists $\epsilon \ge 0$ such that $\hat{\boldsymbol{R}} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0)$, and vice versa. As a result, $\hat{\boldsymbol{R}}$ is feasible to (19), so is $\boldsymbol{R}_0$ to (20).

The theorem below reformulates Problems (19) and (20).

Theorem 1: Problem (19) is equivalent to

$$\min_{\boldsymbol{w} \in \mathcal{W}} \max_{\boldsymbol{R} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0)} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w}, \qquad (22)$$

in the sense of the following: if $(\boldsymbol{w}^*, k^*, \boldsymbol{R}^*)$ solves (19), then $(\boldsymbol{w}^*, \boldsymbol{R}^*)$ solves (22); if $(\boldsymbol{w}^*, \boldsymbol{R}^*)$ solves (22), by constructing

$k^* := \boldsymbol{w}^{*\mathsf{H}} \boldsymbol{R}^* \boldsymbol{w}^* - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0$, then $(\boldsymbol{w}^*, k^*, \boldsymbol{R}^*)$ solves (19). In addition, Problem (20) is equivalent to

$$\min_{\boldsymbol{w} \in \mathcal{W}} \max_{\boldsymbol{R} \in \mathcal{B}_\delta(\hat{\boldsymbol{R}})} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w}. \qquad (23)$$

Proof: See Appendix A. □

Theorem 1 explains why minimax beamforming formulation is robust in the sense of Definition 1, which casts new insights into the robust beamforming community. Specifically, Theorem 1 validates the rationale behind the worst-case optimization (7); see also [9, Eq. (9)], [17, Eq. (46)], and [10, Eq. (10)]. This answers Question Q1) in Subsection II-C.

Considering practicality, we particularly focus on the solution of (23).

Theorem 2: If there exists $\boldsymbol{R}^* \in \mathcal{B}_\delta(\hat{\boldsymbol{R}})$ such that $\boldsymbol{R}^* \succeq \boldsymbol{R}$ for all $\boldsymbol{R} \in \mathcal{B}_\delta(\hat{\boldsymbol{R}})$, then

$$\min_{\boldsymbol{w} \in \mathcal{W}} \max_{\boldsymbol{R} \in \mathcal{B}_\delta(\hat{\boldsymbol{R}})} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} = \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R}^* \boldsymbol{w}.$$

Proof: This can be proven by contradiction. □

Specific applications of Theorem 2 are given as follows, where eigenvalue thresholding, diagonal loading, regularization, and prior-knowledge embedding are shown to be robust; this answers Question Q2) in Subsection II-C.

Example 1 (Eigenvalue Thresholding): Consider

$$\mathcal{B}_\mu(\hat{\boldsymbol{R}}_{\text{thr},\mu}) := \{\boldsymbol{R} : \; \boldsymbol{0} \preceq \boldsymbol{R} \preceq \hat{\boldsymbol{R}}_{\text{thr},\mu}\},$$

where $\hat{\boldsymbol{R}}_{\text{thr},\mu}$ is defined in (6). There exists $\boldsymbol{R}^* := \hat{\boldsymbol{R}}_{\text{thr},\mu}$ such that $\boldsymbol{R}^* \succeq \boldsymbol{R}$ for all $\boldsymbol{R} \in \mathcal{B}_\mu(\hat{\boldsymbol{R}}_{\text{thr},\mu})$. Hence, (23) is particularized into

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{R}}_{\text{thr},\mu} \boldsymbol{w}, \qquad (24)$$

which is the eigenvalue-thresholding beamforming method [8, Eq. (12)], [35]. □

Example 2 (Diagonal Loading): Consider

$$\mathcal{B}_{\epsilon_1}(\hat{\boldsymbol{R}}) := \{\boldsymbol{R} : \; \hat{\boldsymbol{R}} - \epsilon_1 \boldsymbol{I}_N \preceq \boldsymbol{R} \preceq \hat{\boldsymbol{R}} + \epsilon_1 \boldsymbol{I}_N, \; \boldsymbol{R} \succeq \boldsymbol{0}\}, \; (25)$$

for $\epsilon_1 \ge 0$. There exists $\boldsymbol{R}^* := \hat{\boldsymbol{R}} + \epsilon_1 \boldsymbol{I}_N$ such that $\boldsymbol{R}^* \succeq \boldsymbol{R}$ for all $\boldsymbol{R} \in \mathcal{B}_{\epsilon_1}(\hat{\boldsymbol{R}})$. Hence, (23) is particularized into

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} (\hat{\boldsymbol{R}} + \epsilon_1 \boldsymbol{I}_N) \boldsymbol{w}, \qquad (26)$$

which is the diagonal-loading beamforming method. In addition, we consider

$$\mathcal{U}_a := \{\boldsymbol{a} : \; \hat{\boldsymbol{a}} \hat{\boldsymbol{a}}^{\mathsf{H}} - \epsilon_2 \boldsymbol{I}_N \preceq \boldsymbol{a} \boldsymbol{a}^{\mathsf{H}} \preceq \hat{\boldsymbol{a}} \hat{\boldsymbol{a}}^{\mathsf{H}} + \epsilon_2 \boldsymbol{I}_N\},$$

for $\epsilon_2 \ge 0$. Similarly, the constraint in $\mathcal{W}$ [see (11)] can be explicitly expressed as

$$\boldsymbol{w}^{\mathsf{H}} (\hat{\boldsymbol{a}} \hat{\boldsymbol{a}}^{\mathsf{H}} - \epsilon_2 \boldsymbol{I}_N) \boldsymbol{w} \ge 1, \qquad (27)$$

that is, $\mathcal{W} = \{\boldsymbol{w} : \; \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{a}} \hat{\boldsymbol{a}}^{\mathsf{H}} \boldsymbol{w} \ge \epsilon_2 \boldsymbol{w}^{\mathsf{H}} \boldsymbol{w} + 1\}$. This is reminiscent of existing robust beamforming formulation (9) in the literature; however, they are slightly distinct because square roots are involved in the constraint of (9). □

Example 2 gives the diagonal loading method another robustness interpretation against the uncertainty in $\hat{\boldsymbol{R}}$ under the scheme of worst-case optimization, which is technically different from the results in [17, Eq. (46)], [10, Eq. (10)]. This new interpretation brings new insights to the community.

The motivation of constructing $\mathcal{B}_{\epsilon_1}(\hat{\boldsymbol{R}})$ and $\mathcal{U}_a$ as in Example 2 is straightforward: we just assume that the matrix difference $\boldsymbol{R} - \hat{\boldsymbol{R}}$ is two-sided limited, so is the difference $\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} - \hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^{\mathsf{H}}$. In [17, Eq. (51)], (26) is called positive diagonal loading, while (27) is negative diagonal loading. However, [17] obtains (26) and (27) in a way technically different from our treatments as in Theorem 2. Example 2 can be generalized as follows.

Example 3 (Regularization): Consider

$$\mathcal{B}_{\epsilon_1}(\hat{\boldsymbol{R}}) := \{\boldsymbol{R}: \ \hat{\boldsymbol{R}} - \epsilon_1 \boldsymbol{C}_1 \preceq \boldsymbol{R} \preceq \hat{\boldsymbol{R}} + \epsilon_1 \boldsymbol{C}_1, \ \boldsymbol{R} \succeq \boldsymbol{0}\}$$

and

$$\mathcal{U}_a := \{\boldsymbol{a}: \ \hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^{\mathsf{H}} - \epsilon_2 \boldsymbol{C}_2 \preceq \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \preceq \hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^{\mathsf{H}} + \epsilon_2 \boldsymbol{C}_2\},$$

for Hermitian matrices $\boldsymbol{C}_1, \boldsymbol{C}_2 \succeq \boldsymbol{0}$ and $\epsilon_1, \epsilon_2 \geq 0$. Problem (23) is particularized into

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}(\hat{\boldsymbol{R}} + \epsilon_1 \boldsymbol{C}_1)\boldsymbol{w}, \qquad (28)$$

where $\mathcal{W} = \{\boldsymbol{w}: \ \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^{\mathsf{H}}\boldsymbol{w} \geq \epsilon_2 \boldsymbol{w}^{\mathsf{H}}\boldsymbol{C}_2\boldsymbol{w} + 1\}$. This is a general regularized beamforming formulation. □

Example 4 (Prior-Knowledge Embedding): Model (28) is equivalent, in the sense of the same optimal beamformer(s), to

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}(\alpha\hat{\boldsymbol{R}} + \beta\boldsymbol{C}_1)\boldsymbol{w}, \qquad (29)$$

where $\alpha := 1/(1 + \epsilon_1)$, $\beta := \epsilon_1/(1 + \epsilon_1)$, and $\boldsymbol{C}_1$ can be seen as prior knowledge of the unknown $\boldsymbol{R}_0$. This gives the prior-knowledge embedding method [31]. □

In addition to Examples 1-4, another benefit of using Theorem 2 can be seen in robust beamforming based on IPN covariance reconstruction; see Section IV later. In short, we can tailor the uncertainty set $\mathcal{B}_\delta(\hat{\boldsymbol{R}})$ to achieve high-resolution power spectra estimation for better accuracy of IPN covariance reconstruction, especially when interferers are close to the SoI.

D. Globally Distributionally Robust Beamforming

In line with Definition 2 and its motivation, the globally robust beamforming problem can be proposed as

$$\begin{aligned}
(\boldsymbol{w}^*, k^*) = \ & \underset{\boldsymbol{w} \in \mathcal{W}, k}{\arg\min} \ k \\
& \text{s.t.} \quad \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w} - \tau \leq k \cdot d(\boldsymbol{R}, \boldsymbol{R}_0), \ \forall \boldsymbol{R} \in \mathbb{C}^{N \times N}, \\
& \qquad k \geq 0,
\end{aligned}$$
(30)

which finds the globally robust beamformer $\boldsymbol{w}^*$ and the global robustness measure k^*. In real-world operation, (30) can be solved resorting to

$$\begin{aligned}
\min_{\boldsymbol{w} \in \mathcal{W}, k} \ & k \\
\text{s.t.} \quad & \boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w} - \tau \leq k \cdot d(\boldsymbol{R}, \hat{\boldsymbol{R}}), \ \forall \boldsymbol{R} \in \mathbb{C}^{N \times N}, \\
& k \geq 0,
\end{aligned}$$
(31)

for

$$\hat{\boldsymbol{w}}^{\mathsf{H}}\hat{\boldsymbol{R}}\hat{\boldsymbol{w}} \leq \tau < \infty. \qquad (32)$$

Note that $\hat{\boldsymbol{R}}$ is feasible to (30), so is $\boldsymbol{R}_0$ to (31). Considering practicality, we particularly focus on the solution of (31). Let $\boldsymbol{C}$ be a Hermitian invertible weight matrix. We study three cases:

- $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]$.

- $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}, \ \forall \boldsymbol{w} \in \mathbb{C}^N$.
- $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}\boldsymbol{C}^{-1}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}, \ \forall \boldsymbol{w} \in \mathbb{C}^N$.

In all cases, $d(\boldsymbol{R}, \hat{\boldsymbol{R}})$ is a similarity measure between $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$: 1) $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) \geq 0$ for every $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$; 2) $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) = 0$ if and only if $\boldsymbol{R} = \hat{\boldsymbol{R}}$; 3) $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) = d(\hat{\boldsymbol{R}}, \boldsymbol{R})$. We choose $d(\boldsymbol{R}, \hat{\boldsymbol{R}})$ in such ways just for technical tractability compared to, e.g., $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \|\boldsymbol{R} - \hat{\boldsymbol{R}}\|_F = \sqrt{\text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]}$. Under the above three constructions for $d(\boldsymbol{R}, \hat{\boldsymbol{R}})$, the solution to Problem (31) is given below.

Theorem 3: If $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]$, Problem (31) is equivalent, in the sense of the same optimal beamformer(s), to a quartically regularized beamforming problem

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k}\right]\boldsymbol{w} = \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{w} + \frac{1}{4k}(\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w})^2,$$
(33)

where k is chosen to let the above optimal objective value equal to τ. If $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}$, Problem (31) is equivalent, in the sense of the same optimal beamformer(s), to a quadratically regularized (i.e., diagonal-loading) beamforming problem

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\left[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{I}_N\right]\boldsymbol{w} = \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{w} + \frac{1}{4k}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}, \ (34)$$

where k is chosen to let the above optimal objective value equal to τ. If $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}\boldsymbol{C}^{-1}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}$, Problem (31) is equivalent, in the sense of the same optimal beamformer(s), to a general quadratically regularized beamforming problem

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\left[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{C}\right]\boldsymbol{w} = \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{w} + \frac{1}{4k}\boldsymbol{w}^{\mathsf{H}}\boldsymbol{C}\boldsymbol{w}, \ (35)$$

where k is chosen to let the above optimal objective value equal to τ.

Proof: See Appendix B. □

Since $\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathsf{H}}[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k}]\boldsymbol{w}$, $\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathsf{H}}[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{I}_N]\boldsymbol{w}$, and $\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathsf{H}}[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{C}]\boldsymbol{w}$ are continuous and monotonically decreasing in k and they tend to infinity when $k \to 0$, the solutions to the three feasibility problems in Theorem 3 exist. If we require $\tau > \hat{\boldsymbol{w}}^{\mathsf{H}}\hat{\boldsymbol{R}}\hat{\boldsymbol{w}}$, the solutions are guaranteed to be finite (i.e., finite k's exist). Note that when $\tau := \hat{\boldsymbol{w}}^{\mathsf{H}}\hat{\boldsymbol{R}}\hat{\boldsymbol{w}}$, we have $k = \infty$ for all the three cases. An intuitive example of Theorem 3 is given as follows.

Example 5 (Diagonal-Loading Beamforming): Let $\mathcal{U}_a := \{\boldsymbol{a}_0\}$, that is, the steering vector is exactly known; cf. (11). Then the optimal objective of (34) is a function of k, i.e.,

$$\varphi(k) := \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}\left[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{I}_N\right]^{-1}\boldsymbol{a}_0}.$$

Let

$$\tau := \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_0} + t$$

where $t \geq 0$ is a user-design objective-excess parameter; cf. (32). Then, the optimal value of k is given by the zero of the following equation

$$\varphi(k) = \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}\left[\hat{\boldsymbol{R}} + \frac{1}{4k}\boldsymbol{I}_N\right]^{-1}\boldsymbol{a}_0} = \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_0} + t = \tau;$$

that is, k is uniquely determined by t. If $t := 0$, we have $k = \infty$. $\qquad\square$

As seen from Examples 2 and 5 and Theorem 3, diagonal loading is a powerful technique for both locally robust beamforming and globally robust beamforming; this claim is true at least under decent algorithmic constructions. The difference is on which quantity the level of diagonal loading relies: the scale ϵ_1 of the uncertainty set in locally robust beamforming (cf. Example 2) or the prescribed threshold τ in globally robust beamforming (cf. Theorem 3 and Example 5).

E. Regularized Beamforming

Existing literature has well-identified regularized beamforming as an efficient method to achieve robustness against limited snapshot sizes [4], [13], which can also be seen from this article's results in (26), (28), (33), (34), and (35). Another interpretation of regularized beamforming comes with penalizing the array sensitivity against array errors [11], [29]. To be specific, the quantity $\boldsymbol{w}^\mathsf{H}\boldsymbol{w}$ is used as a sensitivity measure of the array; see [11, Section II] and [29, Section III-C] for technical justifications. Intuitively speaking, when the steering vector $\boldsymbol{a}(\theta)$ deviates from its actual value $\boldsymbol{a}_0(\theta)$, the smaller the value of $\boldsymbol{w}^\mathsf{H}\boldsymbol{w}$, the less impact this deviation can cause to beamforming results; note that the derivative of array response pattern $\boldsymbol{w}^\mathsf{H}\boldsymbol{a}$ with respect to steering vector $\boldsymbol{a}$ is $\boldsymbol{w}$; that is, if the norm of $\boldsymbol{w}$ is small, the sensitivity of $\boldsymbol{w}^\mathsf{H}\boldsymbol{a}$ to deviations of $\boldsymbol{a}$ would be also limited.

In this subsection, we introduce the regularization technique from a *different* perspective, i.e., worst-case robustness. We start with the SAA beamforming (14) with the noise injection technique [52, Section IV], [53, Eq. (13)] to achieve robustness. Suppose that the signal $\mathbf{s}$ is contaminated by an unknown zero-mean random error ξ_1 (due to model uncertainties), and the operating array steering vector is contaminated by an unknown zero-mean random error ξ_2 (due to, e.g., calibration and pointing errors). The beamforming problem with noise injection into $\mathbf{s}$ can be formulated as

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^\mathsf{H}\mathbb{E}_{(\mathbf{s}\sim\hat{\mathbb{P}},\ \xi_1\sim\mathbb{P}_1)}[(\mathbf{s}+\xi_1)(\mathbf{s}+\xi_1)^\mathsf{H}]\boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \boldsymbol{w}^\mathsf{H}\mathbb{E}_{\xi_2\sim\mathbb{P}_2}[\xi_2\xi_2^\mathsf{H}]\boldsymbol{w} \geq 1, \tag{36}$$

where we suppose that the distributions of ξ_1 and ξ_2 are $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively; the objective is due to ξ_1-noise injection into $\mathbf{s}$, while the constraint is due to the facts $\hat{\boldsymbol{a}} = \boldsymbol{a}_0 + \xi_2$ and $\boldsymbol{w}^\mathsf{H}\boldsymbol{a}_0\boldsymbol{a}_0^\mathsf{H}\boldsymbol{w} \geq 1$. Assume that $\mathbf{s}$ is uncorrelated with ξ_1. Problem (36) can be rewritten as

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \boldsymbol{R}_1]\boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \boldsymbol{w}^\mathsf{H}\boldsymbol{R}_2\boldsymbol{w} \geq 1, \tag{37}$$

where $\boldsymbol{R}_1 := \mathbb{E}_{\xi_1\sim\mathbb{P}_1}[\xi_1\xi_1^\mathsf{H}]$ and $\boldsymbol{R}_2 := \mathbb{E}_{\xi_2\sim\mathbb{P}_2}[\xi_2\xi_2^\mathsf{H}]$. In practice where we have no knowledge about $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, we can construct the uncertainty sets for $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ respectively as follows:

$$\boldsymbol{R}_1 \in \{\boldsymbol{R}:\ \boldsymbol{0} \preceq \boldsymbol{R} \preceq \epsilon_1\boldsymbol{I}_N\},$$
$$\boldsymbol{R}_2 \in \{\boldsymbol{R}:\ \boldsymbol{0} \preceq \boldsymbol{R} \preceq \epsilon_2\boldsymbol{I}_N\}, \tag{38}$$

for some $\epsilon_1, \epsilon_2 \geq 0$. As a result, the robust counterpart of (37), which optimizes the worst-case performance and guarantees the worst-case feasibility, can be given as

$$\min_{\boldsymbol{w}} \max_{\boldsymbol{R}_1} \quad \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \boldsymbol{R}_1]\boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \max_{\boldsymbol{R}_2}\boldsymbol{w}^\mathsf{H}\boldsymbol{R}_2\boldsymbol{w} \geq 1. \tag{39}$$

The solution to the above problem is given below.

Corollary 1 (of Theorem 2): Problem (39) is equivalent to

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \epsilon_1\boldsymbol{I}_N]\boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \epsilon_2\boldsymbol{w}^\mathsf{H}\boldsymbol{w} \geq 1, \tag{40}$$

which coincides with Example 2. $\qquad\square$

Corollary 1 and (40) reveal an important law in robust beamformer design: to achieve robustness is to add regularization. This law applies to both the objective and the constraint. To be specific, compared with Example 2 and Corollary 1, the insight below is immediate.

Insight 1 (Robustness and Regularization): Consider the SAA beamforming and its robust counterparts. The following two statements are at the core.

- Solving $\min_{\boldsymbol{w}} \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{R}}\boldsymbol{w}$ by $\hat{\boldsymbol{w}}$ at nominal value $\hat{\boldsymbol{R}}$ does not necessarily control true value $\hat{\boldsymbol{w}}^\mathsf{H}\boldsymbol{R}_0\hat{\boldsymbol{w}}$ evaluated at $\boldsymbol{R}_0$. However, solving the regularized problem $\min_{\boldsymbol{w}} \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \epsilon_1\boldsymbol{I}_N]\boldsymbol{w}$ by $\hat{\boldsymbol{w}}'$ can control true value $\hat{\boldsymbol{w}}'^\mathsf{H}\boldsymbol{R}_0\hat{\boldsymbol{w}}'$ evaluated at $\boldsymbol{R}_0$. This is because

$$\boldsymbol{w}^\mathsf{H}\boldsymbol{R}_0\boldsymbol{w} \leq \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \epsilon_1\boldsymbol{I}_N]\boldsymbol{w}, \quad \forall\boldsymbol{w}.$$

  Hence, minimizing the regularized objective means minimizing the upper bound of the true objective function.

- Requiring $\boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} \geq 1$ does not necessarily imply $\boldsymbol{w}^\mathsf{H}\boldsymbol{a}_0\boldsymbol{a}_0^\mathsf{H}\boldsymbol{w} \geq 1$. However, if we require the regularized version, i.e., $\boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \epsilon_2\boldsymbol{w}^\mathsf{H}\boldsymbol{w} \geq 1$, then we can guarantee that $\boldsymbol{w}^\mathsf{H}\boldsymbol{a}_0\boldsymbol{a}_0^\mathsf{H}\boldsymbol{w} \geq 1$ because

$$\boldsymbol{w}^\mathsf{H}\boldsymbol{a}_0\boldsymbol{a}_0^\mathsf{H}\boldsymbol{w} \geq \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \epsilon_2\boldsymbol{w}^\mathsf{H}\boldsymbol{w} \geq 1, \quad \forall\boldsymbol{w}.$$

In short, for both objective and constraint, to achieve robustness is to add regularization. $\qquad\square$

The following corollary solves Problem (39) when the uncertainty sets for $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ take other forms than (38).

Corollary 2 (of Corollary 1): If

$$\boldsymbol{R}_1 \in \{\boldsymbol{R}:\ \boldsymbol{0} \preceq \boldsymbol{R} \preceq \epsilon_1\bar{\boldsymbol{R}}_1\}$$
$$\boldsymbol{R}_2 \in \{\boldsymbol{R}:\ \boldsymbol{0} \preceq \boldsymbol{R} \preceq \epsilon_2\bar{\boldsymbol{R}}_2\}$$

for some $\bar{\boldsymbol{R}}_1, \bar{\boldsymbol{R}}_2 \succeq \boldsymbol{0}$, then Problem (39) is equivalent to

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^\mathsf{H}[\hat{\boldsymbol{R}} + \epsilon_1\bar{\boldsymbol{R}}_1]\boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\mathsf{H}\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^\mathsf{H}\boldsymbol{w} - \epsilon_2\boldsymbol{w}^\mathsf{H}\bar{\boldsymbol{R}}_2\boldsymbol{w} \geq 1, \tag{41}$$

which is a general regularized beamforming formulation than diagonal loading in Corollary 1; see also Example 3. $\qquad\square$

The benefit of using Corollary 2 can be seen in robust beamforming based on IPN covariance reconstruction; see Section IV later. In short, we can design a good regularizer matrix $\bar{\boldsymbol{R}}_1$ to achieve high-resolution power spectra estimation for better accuracy of IPN covariance reconstruction, especially when interferers are close to the SoI.

F. Bayesian-Nonparametric Beamforming

In this subsection, we address the uncertainty in Problem (16), compared to Problem (15), from the perspective of Bayesian nonparametrics.

Let the uncertainty set for the distribution of $\mathbf{s}$ be $\mathcal{P}$; for example, $\mathcal{P}$ can be constructed as a distributional ball $\mathcal{B}_\epsilon(\hat{\mathbb{P}})$ centered at $\hat{\mathbb{P}}$. Since $\hat{\mathbb{P}}$ is not a reliable surrogate of $\mathbb{P}_0$, we assign a probability distribution $\mathbb{Q}$ on the measurable space $(\mathcal{P}, \mathcal{B}_\mathcal{P})$ where $\mathcal{B}_\mathcal{P}$ is the Borel σ-algebra on $\mathcal{P}$. As a result, the Bayesian-nonparametric counterpart of Problem (16) can be given as

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\mathbb{E}_{\mathbb{P}\sim\mathbb{Q}}\mathbb{E}_{\mathbf{s}\sim\mathbb{P}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]\boldsymbol{w}, \tag{42}$$

where $\mathbb{Q}$ is called a second-order probability distribution and $\mathbb{P}$ is a first-order one. Consider the measurable space $(\mathbb{C}^N, \mathcal{B}_{\mathbb{C}^N})$ where $\mathcal{B}_{\mathbb{C}^N}$ denotes the Borel σ-algebra on $\mathbb{C}^N$. For every event $\mathcal{E} \in \mathcal{B}_{\mathbb{C}^N}$, the random probability distribution $\mathbb{P}$ is a realization of $\mathbb{Q}$ and the quantity $\mathbb{P}(\mathcal{E})$ is a random variable taking values on $[0, 1]$; the distribution of the random variable $\mathbb{P}(\mathcal{E})$ is determined by $\mathbb{Q}$. A desired $\mathbb{Q}$ should let $\mathbb{P}$ concentrate around $\mathbb{P}_0$; the more concentrated, the better. Formulation (42) enables assigning a distribution on the covariance matrix and studying the following type of robust beamforming problem

$$\min_{\boldsymbol{w}\in\mathcal{W}} \mathbb{E}_{\mathbf{R}\sim\mathbb{P}_\mathbf{R}}\boldsymbol{w}^{\mathsf{H}}\mathbf{R}\boldsymbol{w},$$

where $\mathbf{R} := \mathbb{E}_{\mathbf{s}\sim\mathbb{P}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]$ and $\mathbb{P}_\mathbf{R}$ is its distribution; for instance, see [54], [55].

In this article, we consider $\mathbb{Q}$ to be a Dirichlet process [56] with base distribution $\bar{\mathbb{P}} \in \mathcal{P}$ and parameter $\alpha \geq 0$. Here, $\bar{\mathbb{P}}$ serves as a prior estimate of $\mathbb{P}_0$ (cf. the sample estimate $\hat{\mathbb{P}}$). Using $\mathbb{Q}$ as a $(\bar{\mathbb{P}}, \alpha)$-Dirichlet process means that for any finite M-partition $(\Xi_1, \Xi_2, \ldots, \Xi_M)$ of $\mathbb{C}^N$, the random vector $(\mathbb{P}(\Xi_1), \mathbb{P}(\Xi_2), \ldots, \mathbb{P}(\Xi_M))$ is distributed according to the Dirichlet distribution whose concentration parameter vector is $(\alpha\bar{\mathbb{P}}(\Xi_1), \alpha\bar{\mathbb{P}}(\Xi_2), \ldots, \alpha\bar{\mathbb{P}}(\Xi_M))$. As a result of [57, Chapter 3], Problem (42) can be equivalently transformed to

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\left[\frac{L}{L+\alpha}\hat{\boldsymbol{R}} + \frac{\alpha}{L+\alpha}\bar{\boldsymbol{R}}\right]\boldsymbol{w}, \tag{43}$$

where $\bar{\boldsymbol{R}} := \mathbb{E}_{\mathbf{s}\sim\bar{\mathbb{P}}}[\mathbf{s}\mathbf{s}^{\mathsf{H}}]$ is the prior knowledge. Intuitively speaking, in (43), we expect the combined probability distribution $[\frac{L}{L+\alpha}\hat{\mathbb{P}} + \frac{\alpha}{L+\alpha}\bar{\mathbb{P}}]$ to be a better estimate of $\mathbb{P}_0$ than $\hat{\mathbb{P}}$ and $\bar{\mathbb{P}}$, so is $[\frac{L}{L+\alpha}\hat{\boldsymbol{R}} + \frac{\alpha}{L+\alpha}\bar{\boldsymbol{R}}]$ to $\boldsymbol{R}_0$ than $\hat{\boldsymbol{R}}$ and $\bar{\boldsymbol{R}}$.

Model (43) gives the prior-knowledge embedding method [31] another rigorous interpretation (cf. Example 4), which offers new insights to the community. In addition, it is equivalent, in the sense of the same optimal beamformer(s), to

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}[\hat{\boldsymbol{R}} + \epsilon\bar{\boldsymbol{R}}]\boldsymbol{w} = \min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{w} + \epsilon\boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{R}}\boldsymbol{w}, \tag{44}$$

where $\epsilon := \alpha/L$. Model (44) is a regularized beamforming problem; cf. Example 3 and Corollary 2.

A direct application of Models (43) and (44) is as follows.

Example 6 (Diagonal Loading): When we do not know the prior knowledge $\bar{\boldsymbol{R}}$, we can set $\bar{\boldsymbol{R}} := \boldsymbol{I}_N$. Problem (43) becomes

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\left[\frac{L}{L+\alpha}\hat{\boldsymbol{R}} + \frac{\alpha}{L+\alpha}\boldsymbol{I}_N\right]\boldsymbol{w},$$

and Problem (44) turns into

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}[\hat{\boldsymbol{R}} + \epsilon\boldsymbol{I}_N]\boldsymbol{w}.$$

The latter is the diagonal-loading method. $\qquad\square$

Since $\alpha \geq 0$ is a user-design parameter in real-world operation, we can write (43) more compactly as

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}[(1 - \beta)\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}}]\boldsymbol{w}, \tag{45}$$

through employing another user-design parameter $\beta \in [0, 1]$. The following insight presents the robustness of the Bayesian-nonparametric beamforming model (45).

Insight 2 (Bayesian-Nonparametric Beamforming): Suppose that under $(\bar{\boldsymbol{R}}, \beta)$, the convex combination $(1 - \beta)\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}}$ is closer to $\boldsymbol{R}_0$ than $\hat{\boldsymbol{R}}$. Then, for $\epsilon_1, \epsilon_2 \geq 0$ such that $(1-\beta)\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}} \in \mathcal{B}_{\epsilon_1}(\boldsymbol{R}_0)$ and $\hat{\boldsymbol{R}} \in \mathcal{B}_{\epsilon_2}(\boldsymbol{R}_0)$, we have $\epsilon_1 \leq \epsilon_2$. As a result, if k_1 solves Problem (19) under $\epsilon := \epsilon_1$ and k_2 solves Problem (19) under $\epsilon := \epsilon_2$, we have $k_1 \leq k_2$. Therefore, compared with the case using $\hat{\boldsymbol{R}}$, employing $(1 - \beta)\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}}$ as a surrogate of $\boldsymbol{R}_0$ would generate a more robust beamformer that has a smaller robustness measure; see Definition 1. $\quad\square$

In addition to Insight 2, another benefit of using Bayesian-nonparametric beamforming method (45) can be seen in robust beamforming based on IPN covariance reconstruction; see Section IV later. In short, we can design the prior matrix $\bar{\boldsymbol{R}}$ to achieve high-resolution power spectra estimation for better accuracy of IPN covariance reconstruction, especially when interferers are close to the SoI.

G. Unified Distributionally Robust Beamforming Framework

Based on the proposals of locally distributionally robust beamforming (20), globally distributionally robust beamforming (31), regularized beamforming (41), and Bayesian-nonparametric beamforming (45), the insight below summarizes a unified distributionally robust beamforming framework.

Insight 3 (Unified Robust Beamforming Framework): A robust beamforming formulation in the sense of Definition 1 should take, or can be transformed into, the following forms:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}[(1 - \beta)\hat{\boldsymbol{R}} + \beta\bar{\boldsymbol{R}}]\boldsymbol{w}$$

or equivalently

$$\min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{w} + \epsilon\boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{R}}\boldsymbol{w},$$

where $\beta \in [0, 1]$ and $\epsilon := \beta/(1 - \beta) \geq 0$; the parameters β, ϵ, and $\bar{\boldsymbol{R}}$ can be determined according to the principles of locally robust beamforming (i.e., Theorems 1 and 2), globally robust beamforming (i.e., Theorem 3), regularized beamforming (i.e., Insight 1), or Bayesian-nonparametric beamforming (i.e., Insight 2). $\quad\square$

For detailed motivations and explanations, also recall Examples 2, 3, 4, 5, and 6, Corollaries 1 and 2, and Model (45).

IV. ROBUSTNESS VIA HIGH-RESOLUTION SPECTRA ESTIMATION AND IPN COVARIANCE RECONSTRUCTION

This section answers Question Q3) in Subsection II-C. Since we aim to estimate the power spectra for IPN covariance

reconstruction, in this section, $\hat{R}$ refers to the snapshot covariance in which the SoI is present. This leads to the Capon (i.e., MPDR) beamforming and Capon power spectra.

The Capon beamforming is believed to be a good-resolution method for power spectra estimation [47]. However, when we move to the robust Capon beamforming regime, the resolution must be sacrificed. This can be straightforwardly seen from the robust counterpart of the Capon constraint (11), i.e., $\min_{\boldsymbol{a}(\theta) \in \mathcal{U}_a} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{a}(\theta) \boldsymbol{a}(\theta)^{\mathsf{H}} \boldsymbol{w} \geq 1$, for $\mathcal{U}_a := \{\boldsymbol{a} : \|\boldsymbol{a} - \boldsymbol{a}_0(\theta)\| \leq \epsilon\}$. To be specific, the robust Capon beamforming maintains a wider beam than the standard Capon beamforming in direction θ to ensure that the array power response $\boldsymbol{w}^{\mathsf{H}} \boldsymbol{a}_0(\theta') \boldsymbol{a}_0(\theta')^{\mathsf{H}} \boldsymbol{w}$ is sufficiently large for all directions θ' close to θ. Note that $\boldsymbol{a}_0(\theta') \in \mathcal{U}_a$ when $|\theta' - \theta|$ is small. In using beamforming for power spectra estimation, the wider the beam, the lower the resolution. Therefore, in the literature, the trade-off between robustness and resolution in beamforming for power spectra estimation largely exists. This dilemma occurs even when the steering vector is exactly known but the SAA-estimate $\hat{R}$ deviates from its true value $\boldsymbol{R}_0$. This is because robustness against array errors technically leads to the diagonal-loading Capon beamforming [6], [7], [11], so does robustness against the uncertainty in $\hat{R}$ (see, e.g., Example 2). Therefore, whenever the diagonal loading technique is applied to achieve robustness, the resolution for power spectra estimation must deteriorate. In this section, we show how to leverage the characteristics of the subspace methods, such as MUSIC, to improve the resolution of power spectra estimation in robust beamforming. This aim is technically realized by designing an appropriate $\bar{R}$ in Insight 3.

We start with the reformulation of the canonical MUSIC method. Let

$$\boldsymbol{U}_0 \boldsymbol{\Sigma}_0 \boldsymbol{U}_0^{\mathsf{H}} := \boldsymbol{R}_0 \qquad (46)$$

denote an eigenvalue decomposition of $\boldsymbol{R}_0$ and eigenvalues in $\boldsymbol{\Sigma}_0$ be sorted in the descending order. We partition $\boldsymbol{U}_0$ and $\boldsymbol{\Sigma}_0$ as follows:

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \boldsymbol{\Sigma}_{0,s} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{0,v} \end{bmatrix}$$

where $\boldsymbol{\Sigma}_{0,s} \in \mathbb{C}^{K \times K}$ and $\boldsymbol{\Sigma}_{0,v} \in \mathbb{C}^{(N-K) \times (N-K)}$, and

$$\boldsymbol{U}_0 = [\boldsymbol{U}_{0,s}, \ \boldsymbol{U}_{0,v}],$$

where the columns of $\boldsymbol{U}_{0,s} \in \mathbb{C}^{N \times K}$ span the signal-plus-interference subspace and the columns of $\boldsymbol{U}_{0,v} \in \mathbb{C}^{N \times (N-K)}$ span the noise subspace. Let

$$\hat{\boldsymbol{U}} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{U}}^{\mathsf{H}} := \hat{\boldsymbol{R}} \qquad (47)$$

denote an eigenvalue decomposition of $\hat{\boldsymbol{R}}$ and eigenvalues in $\hat{\boldsymbol{\Sigma}}$ be sorted in the descending order. Similarly, $\hat{\boldsymbol{\Sigma}}$ is partitioned into $\hat{\boldsymbol{\Sigma}}_s$ and $\hat{\boldsymbol{\Sigma}}_v$, so is $\hat{\boldsymbol{U}}$ into $\hat{\boldsymbol{U}}_s$ and $\hat{\boldsymbol{U}}_v$.

The MUSIC pseudo-spectra are defined as $P_{\mathrm{M}}(\theta) := 1/[\boldsymbol{a}_0^{\mathsf{H}}(\theta) \boldsymbol{U}_{0,v} \boldsymbol{U}_{0,v}^{\mathsf{H}} \boldsymbol{a}_0(\theta)]$, that is,

$$P_{\mathrm{M}}(\theta) = \frac{\sigma_n^{-2}}{\boldsymbol{a}_0^{\mathsf{H}}(\theta) [\boldsymbol{U}_{0,s}, \ \boldsymbol{U}_{0,v}] \begin{bmatrix} \boldsymbol{0} & \\ & \boldsymbol{\Sigma}_{0,v}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{0,s}^{\mathsf{H}} \\ \boldsymbol{U}_{0,v}^{\mathsf{H}} \end{bmatrix} \boldsymbol{a}_0(\theta)}.$$

Note that $\boldsymbol{\Sigma}_{0,v} = \sigma_n^2 \boldsymbol{I}_{N-K}$ where σ_n^2 denotes the power of channel noise. Because the constant σ_n^{-2} does not influence the power pattern, we can ignore it and modify the MUSIC pseudo-spectra to

$$P_{\mathrm{M}}(\theta) = \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}(\theta) [\boldsymbol{U}_{0,s}, \ \boldsymbol{U}_{0,v}] \begin{bmatrix} \boldsymbol{0} & \\ & \boldsymbol{\Sigma}_{0,v}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{0,s}^{\mathsf{H}} \\ \boldsymbol{U}_{0,v}^{\mathsf{H}} \end{bmatrix} \boldsymbol{a}_0(\theta)}. \qquad (48)$$

In contrast, the Capon power spectra, which solve Problem (5), are given by $P(\theta) = 1/[\boldsymbol{a}_0^{\mathsf{H}}(\theta) \boldsymbol{R}_0^{-1} \boldsymbol{a}_0(\theta)]$, that is,

$$P(\theta) = \frac{1}{\boldsymbol{a}_0^{\mathsf{H}}(\theta) [\boldsymbol{U}_{0,s}, \ \boldsymbol{U}_{0,v}] \begin{bmatrix} \boldsymbol{\Sigma}_{0,s}^{-1} & \\ & \boldsymbol{\Sigma}_{0,v}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{0,s}^{\mathsf{H}} \\ \boldsymbol{U}_{0,v}^{\mathsf{H}} \end{bmatrix} \boldsymbol{a}_0(\theta)}. \qquad (49)$$

Comparing MUSIC pseudo-spectra in (48) with Capon spectra in (49), we have the following key observation.

Insight 4: In Capon beamforming and its variants, from the perspective of algorithm design, the key to high-resolution power spectra estimation is to downweight the signal-plus-interference subspace component $\boldsymbol{a}_0^{\mathsf{H}}(\theta) \boldsymbol{U}_{0,s} \boldsymbol{\Sigma}_{0,s}^{-1} \boldsymbol{U}_{0,s}^{\mathsf{H}} \boldsymbol{a}_0(\theta)$. One way is to modify $\boldsymbol{\Sigma}_{0,s}^{-1}$ in (49) to $\gamma \boldsymbol{\Sigma}_{0,s}^{-1}$ where $0 \leq \gamma \leq 1$ is a small value. Another choice is to modify $\boldsymbol{\Sigma}_{0,s}^{-1}$ in (49) to $[\boldsymbol{\Sigma}_{0,s} + \gamma \boldsymbol{I}_K]^{-1}$ where $\gamma \geq 0$ is a large value. $\square$

A. Robust Beamforming for High-Resolution Power Spectra Estimation

Combining the principles of robustness in Insight 3 and high resolution in Insight 4, the robust beamforming methods for high-resolution power spectra estimation can be put forward. Let

$$\boldsymbol{\Gamma} := \hat{\boldsymbol{U}} \begin{bmatrix} \delta_1 \boldsymbol{I}_{K'} & \\ & \delta_2 \boldsymbol{I}_{N-K'} \end{bmatrix} \hat{\boldsymbol{U}}^{\mathsf{H}}, \qquad (50)$$

for $\delta_1, \delta_2 \geq 0$ and $K \leq K' \leq N - 1$.

Consider the robust beamforming problem

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} (\hat{\boldsymbol{R}} + \boldsymbol{\Gamma}) \boldsymbol{w}. \qquad (51)$$

The above formulation can be seen as an **unbalanced diagonal loading** (UDL) method, which generalizes the conventional balanced diagonal loading (BDL) method where $\delta_1 = \delta_2$. Note that for the purpose of high-resolution power spectra estimation, K' can be any integer value in $[K, N-1]$; see [47, Section V], [58, Section 4.5] for explanations and justifications. Note also that for the purpose of robustness, we only expect $\boldsymbol{\Gamma}$ to be positive semi-definite such that $\boldsymbol{w}^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w} \leq \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{R}} \boldsymbol{w} + \boldsymbol{w}^{\mathsf{H}} \boldsymbol{\Gamma} \boldsymbol{w}$, for every $\boldsymbol{w}$, regardless of the value of K'; cf. Insights 1 and 3.

Suppose $\mathcal{U}_a := \{\boldsymbol{a}_0(\theta)\}$; that is, the steering vector is exactly known; cf. (11). We have the following observations:

- If $\delta_1 = \delta_2 = 0$, Model (51) gives the standard Capon beamforming method.
- If $\delta_1 = \delta_2 \neq 0$, Model (51) gives the conventional BDL method; i.e., the two loading levels are the same.
- If $\delta_1 \neq \delta_2$, Model (51) presents a UDL method; i.e., the two loading levels are different.
- For any $K' \in [K, N-1]$, if $\delta_1 \to \infty$, Model (51) generates the high-resolution subspace pseudo-spectra

[47, Section 5]. When $K' = K$, this subspace pseudo-spectra become the MUSIC pseudo-spectra (48). When $K' = N - 1$, this subspace pseudo-spectra become the Pisarenko pseudo-spectra [58, p. 162].

Motivated by the preceding observations, we propose the following robust beamforming for high-resolution power spectra estimation.

Method 1: Solve robust Capon beamforming (51) with $\mathbf{\Gamma}$ in (50). To achieve high resolution for power spectral estimation, let $\delta_1 \gg \delta_2 \geq 0$; to achieve robustness against the uncertainty in $\hat{\mathbf{R}}$, let $\delta_1, \delta_2 > 0$. □

B. Robust Beamforming via IPN Covariance Reconstruction

Motivated by Insight 3 and Method 1, this subsection summarizes the algorithm for robust beamforming based on high-resolution power spectra estimation and IPN covariance reconstruction; see Algorithm 1. In practice, since δ_1 and δ_2 in (50) are still user-design parameters, the employment of tradeoff parameters ϵ_1 in (41) and β in (45) are dispensable. Therefore, in practice, we can set $\epsilon_1 := 1$ and $\beta := 0.5$.

Algorithm 1 Robust beamforming based on high-resolution power spectra estimation and IPN covariance reconstruction

---

Basics: Insight 3, Insight 4, and Method 1.
Notes: In defining $\mathbf{\Gamma}$, to ensure high resolution and sufficient robustness, let $\delta_1 \gg \delta_2 > 0$. In evaluating (10), the integration is approximated using a summation; cf. [40, Eq. (20)].

Input: $\hat{\mathbf{R}}$, $\mathbf{\Gamma}$, $\hat{\mathbf{a}}$, ϵ
  // Step 1: High-Resolution Power Spectral Estimation
    $P(\theta) = \frac{1}{\hat{\mathbf{a}}^\mathsf{H}(\theta)[\hat{\mathbf{R}}+\mathbf{\Gamma}]^{-1}\hat{\mathbf{a}}(\theta)}$
  // Step 2: IPN Covariance Reconstruction
    Use $P(\theta)$, $\hat{\mathbf{a}}$, and (10) to reconstruct $\hat{\mathbf{R}}_{i+n}$
  // Step 3: Robust Beamforming using Diagonal Loading
    $\mathbf{w}^* = (\hat{\mathbf{R}}_{i+n} + \epsilon\mathbf{I}_N)^{-1}\hat{\mathbf{a}}/[\hat{\mathbf{a}}^\mathsf{H}(\hat{\mathbf{R}}_{i+n} + \epsilon\mathbf{I}_N)^{-1}\hat{\mathbf{a}}]$
Output: Robust beamformer $\mathbf{w}^*$

In Step 1, the unbalanced diagonal loading operation does not alter the main characteristics of the power spectrum (e.g., the locations of peaks), since it perturbs only the eigenvalues while preserving the eigenvectors; cf. (50). Although eigenvalues affect the shape and amplitude scaling of the spectrum (thereby allowing control over resolution), the angular selectivity in a power spectrum is primarily governed by the eigenvectors [46]; for empirical illustration, see Fig. 3 in the experiment section. Therefore, the directions of interferences and their relative power levels, as reflected in the spectrum, can be effectively identified. As a result, the IPN covariance reconstruction in Step 2 remains valid and captures the essential statistical information of the IPN. Consequently, in Step 3, the IPN components present in the snapshots can be effectively suppressed.

The computational complexity of Algorithm 1 is as follows.
Remark 1 (Computationally Complexity): The computational complexity of Algorithm 1 is comparable to the method in [39]. The differences between the two approaches are two-fold: first, when estimating the power spectra, Algorithm 1

uses $\hat{\mathbf{R}}+\mathbf{\Gamma}$, instead of $\hat{\mathbf{R}}$, to enhance resolution and robustness; second, diagonal loading, i.e., $\hat{\mathbf{R}}_{i+n} + \epsilon\mathbf{I}_N$, is applied to ensure the robustness against the uncertainty in $\hat{\mathbf{R}}_{i+n}$. The two differences lead to additional $\mathcal{O}(N^2)$ FLOPs (floating-point operations). Hence, the main computational burden of Algorithm 1 is at evaluating the integral in (10), which admits $\mathcal{O}(N^2S)$ FLOPs; S is the number of sampling points in $\bar{\Theta}$; NB: $S \gg N$. As a result, the overall complexity of Algorithm 1 is $\mathcal{O}(N^2S)$, the same as the method in [39]. □

In addition, the following remark is practically instructive.
Remark 2 (Design $\mathcal{W}$): Technically, addressing either the uncertainty in the assumed steering vector or that in the estimated IPN covariance leads to the regularization operation (e.g., diagonal loading method); see [6], [7], [11] and Insight 3. Therefore, in real-world operation, there is no need to consider the uncertainty in the steering vector; that is, the uncertainty set for the steering vector contains only a singleton:

$$\mathcal{U}_a(\theta) := \{\hat{\mathbf{a}}(\theta)\}, \qquad \forall\theta; \qquad (52)$$

see (11). This trick greatly reduces the operational complexity of Algorithm 1. However, the robustness against the uncertainties in the assumed steering vector and the estimated IPN covariance remains. If general $\mathcal{U}_a$ are preferred, refer to handling methods in, e.g., [6], [8], [10], [50], [51]. □

V. EXPERIMENTS

In this section, we consider a uniform linear array with $N = 10$ and half-wavelength spacing [6], [39]. We show how the unbalanced diagonal loading trick in Method 1 benefits the suppression of closely located interferers. Our method complements the state of the art because existing studies assume that interferers are sufficiently separated in their experiments; see, e.g., [6], [39], [43]. We suppose that the SoI is at $\theta_1 = -30°$, and two **strong** interferers with interference-to-signal ratio (ISR) of 10dB are located at $-22°$ and $30°$. In IPN covariance reconstruction, we let the integration interval in (10) be $\bar{\Theta} := [-90°, -35°] \cup [-25°, 90°]$ (cf. [39]); that is, the interferer at $\theta_2 = -22°$ is close to the boundary of the uncertainty region $[-35°, -25°]$ of the SoI. Throughout the experiments, we let $\hat{\mathbf{a}}(\theta) := \mathbf{a}_0(\theta) + \mathbf{\Delta}$ where $\mathbf{\Delta}$ is a Gaussian random vector with covariance $0.01\mathbf{I}_N$; for every Monte-Carlo episode, this $\mathbf{\Delta}$ changes its value. The output SINR performance of each beamformer is averaged over 500 Monte-Carlo simulations. All the source data and codes are available online at GitHub: https://github.com/Spratm-Asleaf/Beamforming-UDL.

A. Experimental Verification of Robustness Principles

In this subsection, we provide empirical verifications of the two robustness principles (i.e., Bayesian and Regularization) in Insight 3. Without loss of generality for demonstration, the MPDR (i.e., Capon) beamforming scheme is employed. Accordingly, we refer to the first robustness formulation as Capon-Bayesian, while the second as Capon-Regularization. We sweep the value of β from 0 to 1 with step size 0.2, and fix $\epsilon = 0.25$ so that, for $\beta = 0.2$, we have $\epsilon = 0.25 = 0.2/0.8 = \beta/(1-\beta)$. The output SINR is plotted against input signal-to-noise ratio (SNR). The number L of snapshots is 30.

(Other values of ϵ and L do not change the empirical claims in this subsection.) For performance comparison, the following two methods are also implemented:

- Optimal: The optimal beamformer where the true IPN covariance matrix $\boldsymbol{R}_{i+n}$ and the true steering vector $\boldsymbol{a}_0(\theta)$ are exactly known. Note that, in real-world operations, this method is not applicable.
- Capon: The Capon beamformer with estimated snapshot covariance $\hat{\boldsymbol{R}}$ and assumed steering vector $\hat{\boldsymbol{a}}(\theta)$.

First, we assume that no effective prior information about the true snapshot covariance $\boldsymbol{R}_0$ is available, so we let $\bar{\boldsymbol{R}} := \boldsymbol{I}_N$. Consequently, the second principle in Insight 3, i.e., the regularization formulation $\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{R}} \boldsymbol{w} + \epsilon \boldsymbol{w}^{\mathsf{H}} \bar{\boldsymbol{R}} \boldsymbol{w}$, specifies the diagonal loading method with loading level $\epsilon = 0.25$, which provides the robustness against steering-vector uncertainties and background white noises. The experimental results are shown in Fig. 1, for which we have the following remarks:

- In the Bayesian non-parametric principle, convexly combining the empirical estimate $\hat{\boldsymbol{R}}$ with the prior guess $\bar{\boldsymbol{R}}$ can potentially improve the robustness.
- In the regularization principle, employing the regularizer $\bar{\boldsymbol{R}}$ can also potentially improve the robustness.
- When $\beta = 0.0$, Capon-Bayesian reduces to the Capon method; when $\beta = 0.2$, Capon-Bayesian is equivalent to Capon-Regularization because $\epsilon = 0.25 = \beta/(1-\beta)$; when $\beta = 1.0$, Capon-Bayesian reduces to the conventional non-adaptive Bartlett beamformer.
- When β is well tuned (e.g., $0.4, 0.6, 0.8$), Capon-Bayesian outperforms Capon-Regularization with $\epsilon = 0.25$.

Second, we suppose that the two interferers are known to be located within $[-24, -20]$ and $[28, 32]$, respectively. This prior information can be used to construct $\bar{\boldsymbol{R}}$. To be specific, letting $\bar{\Theta}_1$ and $\bar{\Theta}_2$ be sampling subsets of $[-24, -20]$ and $[28, 32]$, respectively, with step size of 0.1 degrees, we have

$$\bar{\boldsymbol{R}} := \sum_{\theta_i \in \bar{\Theta}_1} \hat{\boldsymbol{a}}(\theta_i)\hat{\boldsymbol{a}}^{\mathsf{H}}(\theta_j) + \sum_{\theta_j \in \bar{\Theta}_2} \hat{\boldsymbol{a}}(\theta_j)\hat{\boldsymbol{a}}^{\mathsf{H}}(\theta_i) + \boldsymbol{I}_N. \quad (53)$$

The first two terms account for interference suppression, whereas the identity matrix $\boldsymbol{I}_N$ provides the robustness against steering-vector uncertainties and background white noises (cf. diagonal loading). Tradeoff coefficients for the three terms can be introduced; however, for demonstration purposes, it suffices to use (53) as an example. To show performance improvement when effective prior information is incorporated, in the Bayesian non-parametric principle, we let $\bar{\boldsymbol{R}}$ be defined in (53), while in the regularization principle, we keep $\bar{\boldsymbol{R}} = \boldsymbol{I}_N$ unchanged (as in Fig. 1). The experimental results are shown in Fig. 2, for which we have the following remarks:

- In the Bayesian non-parametric formulation, effective prior information indeed helps improve the ability of interference suppression.
- In the Bayesian non-parametric formulation, overly emphasizing the prior information (with large β), on the contrary, can deteriorate the performance because the resulting beamformer loses adaptivity to snapshots. For example, see performance degradation in Fig. 2(f) when input SNR is small.
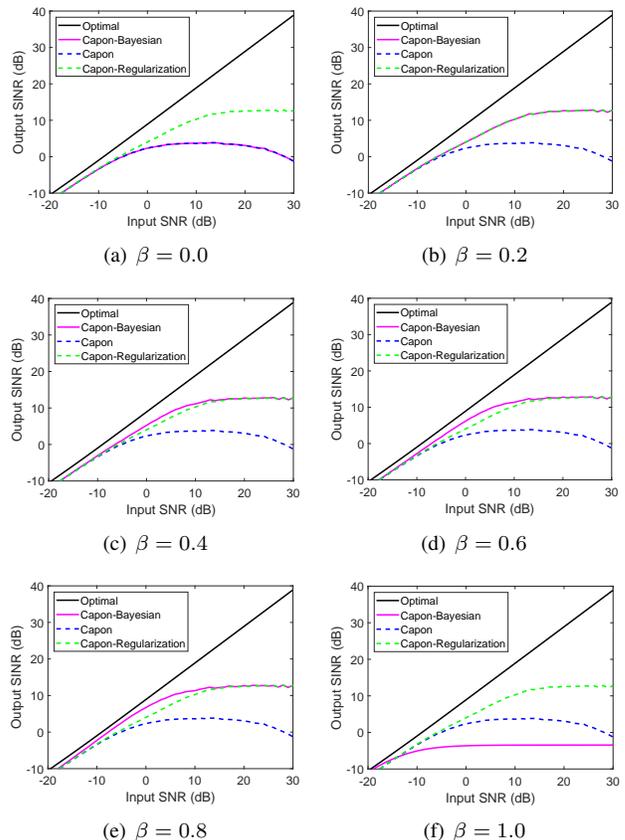

Fig. 1. Empirical verifications of the robustness principles in Insight 3 when no effective prior information about interferers is assumed.

In summary, both robustness formulations in Insight 3 can provide significant robustness, especially when effective prior information can be given. Although the two formulations are developed from different motivations and techniques, given the same $\bar{\boldsymbol{R}}$, they are equivalent if $\epsilon = \beta/(1-\beta)$. Due to this equivalence, we do not explicitly differentiate the two formulations in later experiments; instead, we just leverage the regularization formulation.

B. Experimental Verification of Algorithm 1

In this subsection, we study the empirical performance of the proposed Algorithm 1. In addition to the previously defined Optimal and Capon, the following state-of-the-art methods are implemented in the experiments for performance comparison:

- IPN: Robust beamforming based on IPN covariance reconstruction [39].
- IPN-DL: Diagonally-loaded robust beamforming based on IPN covariance reconstruction. Note that the estimation of IPN covariance can never be exact in practice. Hence, robust methods such as diagonal loading have the potential to further improve the performance of the approach in [39].
- IPN-UDL: Algorithm 1 of this article. IPN-DL is a special case of IPN-UDL for $\delta_1 = \delta_2 = 0$.
- IPN-MaxEnt: Robust beamforming based on maximum-entropy (MaxEnt) spectra estimation and IPN covariance

(a) $\beta = 0.0$

(b) $\beta = 0.2$

(c) $\beta = 0.4$

(d) $\beta = 0.6$

(e) $\beta = 0.8$

(f) $\beta = 1.0$
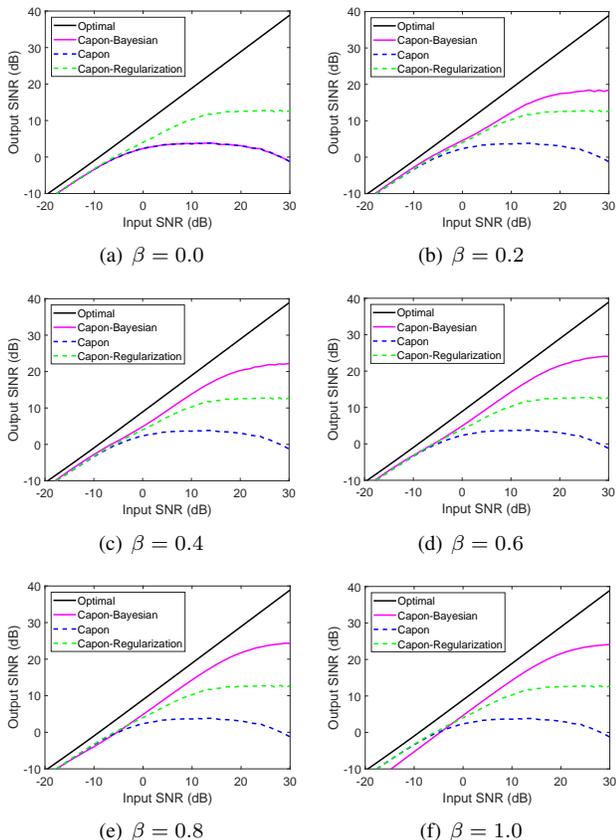
Fig. 2. Empirical verifications of the robustness principles in Insight 3 when effective prior location information about interferers is assumed.

reconstruction [40]. Although the MaxEnt spectra tend to have better resolution than the Capon spectra, the biases of the MaxEnt spectra can be large [46, Fig. 4]. Hence, the performance advantage of the IPN-MaxEnt method [40] over the IPN method [39] may significantly vary across different scenarios or different Monte-Carlo trials.

- Capon-DL: Diagonally-loaded Capon beamformer.

The first four methods are under the MVDR beamforming scheme, while the last method is under the MPDR beamforming scheme. We do not study other existing methods because they are greatly dominated by IPN [39] and IPN-MaxEnt [40]. Capon and Capon-DL are kept just for the illustration of the theoretical analyses in this article.

1) Power Spectra Estimation: The power spectra estimated by Capon beamforming, Capon beamforming with diagonal loading, and Capon beamforming with unbalanced diagonal loading are shown in Fig. 3. Note that in evaluating the resolution of a power spectra estimation method, all emitting signals should have the same energy; that is, the interference-to-signal ratio should be 0dB [47, p. 1020]. Fig. 3 suggests that the diagonal loading operation indeed widens the beams, and hence, deteriorates the resolution. This experimental observation agrees with the theoretical analyses in the beginning of Section IV. However, the unbalanced diagonal loading trick can improve the resolution in power spectra estimation, aligning with the motivation in Insight 4 and Method 1. Note that the larger the degree of unbalancedness is (i.e., the larger

the value of δ_1 when fixing δ_2), the better the resolution is.


(a) $\delta_1 = 2$

(b) $\delta_1 = 6$
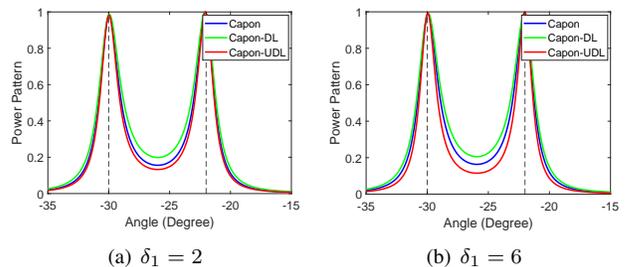
Fig. 3. Power spectra of Capon beamforming, Capon beamforming with diagonal loading (DL), and Capon beamforming with Unbalanced DL (UDL). In this scenario, ISR = 0dB and SNR = 25dB; the number of snapshots is 30; in Capon-DL, the DL parameter is $\epsilon = 0.01$; in Capon-UDL, $\delta_2 = 0.0$ (cf. Insight 4).

2) Output SINR: To compare output SINR performances, all beamforming methods are empirically tuned to achieve their best operational performances in the investigated scenarios. To be specific, in IPN-DL, the diagonal loading parameter is $\epsilon = 0.01$; in IPN-UDL, $K' = K = 3$, $\delta_1 = 10$, and $\delta_2 = \epsilon = 0.01$ (NB: large δ_1 for high-resolution and nonzero δ_2 for robustness; see Method 1); in Capon-DL, the diagonal loading parameter is $\epsilon = 0.1$.

The output SINR performance of beamformers is shown in Fig. 4. When plotting against the input SNR, the number of snapshots is set to 30 or 80; when plotting against the number of snapshots, the input SNR is set to 10dB or 25dB.

Fig. 4 outlines the following observations.

- For both IPN and Capon, diagonal loading has the potential to improve their robustness, especially in the regime of small to moderate SNRs. This is because the IPN covariance in the MVDR beamforming and the snapshot covariance in the MPDR beamforming can never be exactly estimated. Therefore, according to the analyses in Section III (e.g., Insights 1, 2, and 3), robust beamforming methods tend to outperform their nominal counterparts.

- IPN-UDL significantly outperforms the other beamformers due to its high-resolution and low-bias in power spectra estimation; cf. Fig. 3.

3) Supplementary Experiments: In the experiment associated with Fig. 4, we have assumed that $K' = K = 3$. Below, we relax this assumption and admit that K' may not be exactly known. We set $K' = 5$ or $K' = 7$. The corresponding results are given in Fig. 5. As we can see, the proposed IPN-UDL method is not sensitive to the value of K', as long as it is reasonably set in practice. However, when $K' = 7$, the performance indeed slightly degrades compared to that for $K' = 5$ and $K' = K = 3$. This experimental observation agrees with our theoretical analyses in Subsection IV-A.

When the interference-to-signal ratio is 0dB and the locations of interferers are at $\theta_2 = 0°$ and $\theta_3 = 30°$, the output SINR performance against the input SNR is shown in Fig. 6. In this case, the interferers are far away from the SoI and the interference energies are equal to that of the SoI. Therefore, as reported in [39], the IPN method is indeed a powerful tool that can achieve nearly optimal performance in both low and high SNR regimes.
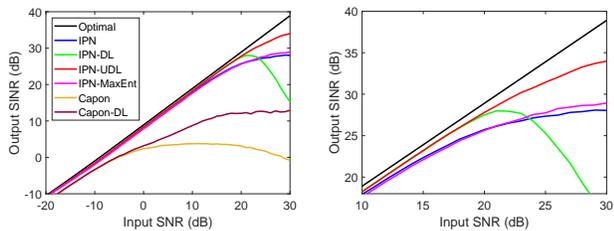
(a) Against SNR (Snapshot = 30) (b) Closeup of (a)
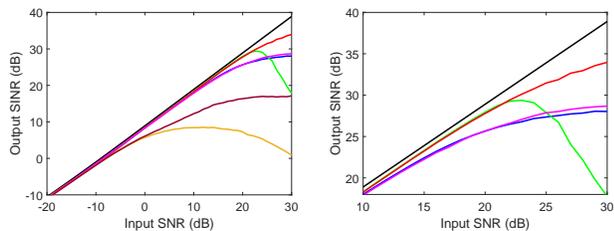
(c) Against SNR (Snapshot = 80) (d) Closeup of (c)
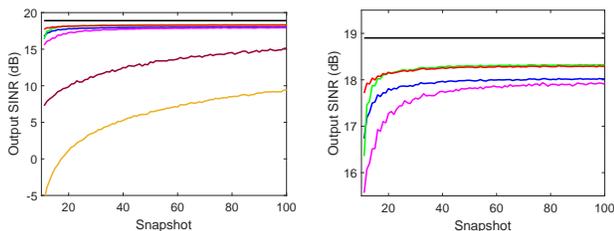
(e) Against Snapshot (SNR = 10dB) (f) Closeup of (e)
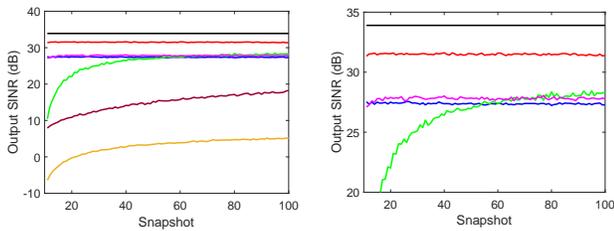
(g) Against Snapshot (SNR = 25dB) (h) Closeup of (g)

Fig. 4. The output SINR performance of beamformers against input SNR and snapshots. In (a), the number of snapshots is 30; in (c), the number of snapshots is 80; in (e), the SNR is 10dB; in (g), the SNR is 25dB.
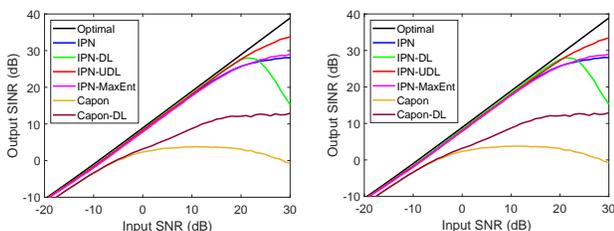

(a) $K' = 5$ (b) $K' = 7$

Fig. 5. The output SINR performance of beamformers against input SNR when $K' = 5$ and $K' = 7$. The number of snapshots is 30.
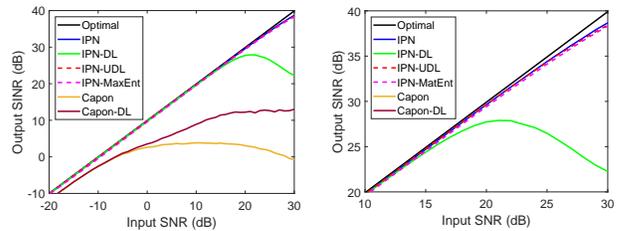
(a) Far and Weak Interferers (b) Closeup of (a)

Fig. 6. The output SINR performance of beamformers against input SNR (far and weak interferers). The ISR is 0dB and the number of snapshots is 30.

VI. CONCLUSIONS

To solve the problems in Subsection I-B and answer the questions in Subsection II-C, this article comprehensively studies distributionally robust beamforming, including the conceptual system, theoretical analysis, and algorithmic design. By addressing uncertainties in snapshot and IPN covariance matrices, a unified framework for distributionally robust beamforming that generalizes several existing approaches (e.g., diagonal loading, eigenvalue thresholding, and prior-knowledge embedding) is presented (see Insight 3), which draws insights from four technical approaches: locally robust, globally robust, regularized, and Bayesian-nonparametric beamforming. Our theoretical analyses demonstrate the equivalence among the four methods, and show that integrating them with subspace methods using the unbalanced diagonal-loading trick (50) can enhance the resolution of power spectra (and DoA) estimation. Such high resolution can improve the reconstruction accuracy of IPN covariances, and hence, improve the SINR performance when interferers are located close to the SoI.

APPENDIX A
PROOF OF THEOREM 1

Proof: First, Problem (19) is equivalent to

$$\min_{\boldsymbol{w} \in \mathcal{W}} \min_{k} \quad k$$
$$\text{s.t.} \quad \max_{\boldsymbol{R} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0)} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0 \leq k,$$
$$k \geq 0.$$

By eliminating k, the above display is equivalent to

$$\min_{\boldsymbol{w} \in \mathcal{W}} \max_{\boldsymbol{R} \in \mathcal{B}_\epsilon(\boldsymbol{R}_0)} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0,$$

which is further equivalent to (22). Note that $\max_{\boldsymbol{R}} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - \boldsymbol{w}_0^{\mathsf{H}} \boldsymbol{R}_0 \boldsymbol{w}_0 \geq 0$ because $\boldsymbol{R}_0 \in \mathcal{B}_\epsilon(\boldsymbol{R}_0)$. The second statement can be similarly proven. $\square$

APPENDIX B
PROOF OF THEOREM 3

Proof: If $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]$, Problem (31) can be written as

$$\min_{\boldsymbol{w} \in \mathcal{W}, k} \quad k$$
$$\text{s.t.} \quad \max_{\boldsymbol{R} \in \mathbb{C}^{N \times N}} \boldsymbol{w}^{\mathsf{H}} \boldsymbol{R} \boldsymbol{w} - k \, \text{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}] \leq \tau.$$
$$k \geq 0.$$

In terms of $\boldsymbol{R}$, this is a quadratic convex and unconstrained optimization, and therefore, the first-order optimality condition $2k(\boldsymbol{R} - \hat{\boldsymbol{R}}) = \boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}$ gives the globally optimal solution $\boldsymbol{R} = \hat{\boldsymbol{R}} + \boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}/(2k)$. Hence, the above optimization is equivalent to

$$\min_{\boldsymbol{w} \in \mathcal{W}, k} \quad k$$
$$\text{s.t.} \quad \boldsymbol{w}^{\mathsf{H}} \left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k} \right] \boldsymbol{w} \leq \tau$$
$$k \geq 0.$$

Because $\boldsymbol{w}^{\mathsf{H}} \left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k} \right] \boldsymbol{w}$ is continuous and monotonically decreasing in k for every $\boldsymbol{w} \in \mathcal{W}$, the above optimization is further equivalent to

$$\min_{k} \quad k$$
$$\text{s.t.} \quad \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k} \right] \boldsymbol{w} \leq \tau$$
$$k \geq 0.$$

Since $\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k} \right] \boldsymbol{w}$ is also continuous and monotonically decreasing in k and it tends to infinity when $k \to 0$, the above optimization is equivalent to

$$\text{find} \quad k$$
$$\text{s.t.} \quad \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^{\mathsf{H}} \left[\hat{\boldsymbol{R}} + \frac{\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}}{4k} \right] \boldsymbol{w} = \tau$$
$$k \geq 0.$$

This proves the first part of the theorem. The second part is technically similar; just note that

$$d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \operatorname{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}$$
$$= \boldsymbol{w}^{\mathsf{H}}\boldsymbol{w} \operatorname{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}[\boldsymbol{R} - \hat{\boldsymbol{R}}],$$

and therefore, the first-order optimality condition $2k\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}(\boldsymbol{R} - \hat{\boldsymbol{R}}) = \boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}$ gives the globally optimal solution

$$\boldsymbol{R} = \hat{\boldsymbol{R}} + \boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}/(2k\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}).$$

The third part is also technically similar; just note that

$$d(\boldsymbol{R}, \hat{\boldsymbol{R}}) := \boldsymbol{w}^{\mathsf{H}} \operatorname{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}\boldsymbol{C}^{-1}[\boldsymbol{R} - \hat{\boldsymbol{R}}]\boldsymbol{w}$$
$$= \boldsymbol{w}^{\mathsf{H}}\boldsymbol{w} \operatorname{Tr}[\boldsymbol{R} - \hat{\boldsymbol{R}}]^{\mathsf{H}}\boldsymbol{C}^{-1}[\boldsymbol{R} - \hat{\boldsymbol{R}}],$$

and therefore, the first-order optimality condition $2k\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}\boldsymbol{C}^{-1}(\boldsymbol{R} - \hat{\boldsymbol{R}}) = \boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}$ gives the globally optimal solution

$$\boldsymbol{R} = \hat{\boldsymbol{R}} + \boldsymbol{C}\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}/(2k\boldsymbol{w}^{\mathsf{H}}\boldsymbol{w}).$$

This completes the proof. $\qquad \square$

REFERENCES

[1] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation theory*. John Wiley & Sons, 2002.

[2] J. Li and P. Stoica, *Robust Adaptive Beamforming*. John Wiley & Sons, 2006.

[3] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. John Wiley & Sons, 2017.

[4] A. M. Elbir, K. V. Mishra, S. A. Vorobyov, and R. W. Heath, "Twenty-five years of advances in beamforming: From convex and nonconvex optimization to learning techniques," *IEEE Signal Process. Mag.*, vol. 40, no. 4, pp. 118–131, 2023.

[5] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of MVDR and MPDR beamformers," in *IEEE 26th Convention of Electrical and Electronics Engineers in Israel*. IEEE, 2010, pp. 416–420.

[6] S. A. Vorobyov, A. B. Gershman, and 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, 2003.

[7] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, 2003.

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

[9] 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. 4, pp. 1539–1547, 2008.

[10] Y. Huang, H. Fu, S. A. Vorobyov, and Z.-Q. Luo, "Robust adaptive beamforming via worst-case SINR maximization with nonconvex uncertainty sets," *IEEE Trans. Signal Process.*, vol. 71, pp. 218–232, 2023.

[11] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, 1987.

[12] J. R. Guerci, "Theory and application of covariance matrix tapers for robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 977–985, 1999.

[13] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Process.*, vol. 93, no. 12, pp. 3264–3277, 2013.

[14] A. Khabbazibasmenj, S. A. Vorobyov, and A. Hassanien, "Robust adaptive beamforming based on steering vector estimation with as little as possible prior information," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2974–2987, 2012.

[15] A. Steele, "Comparison of directional and derivative constraints for beamformers subject to multiple linear constraints," in *IEE Proceedings H (Microwaves, Optics and Antennas)*, vol. 130, no. 1. IET, 1983, pp. 41–45.

[16] M. Er and A. Cantoni, "An alternative formulation for an optimum beamformer with robustness capability," in *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 132, no. 6. IET, 1985, pp. 447–460.

[17] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, 2003.

[18] M. Zatman and J. R. Guerci, "Comments on" theory and application of covariance matrix tapers for robust adaptive beamforming"[with reply]," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1796–1800, 2000.

[19] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, 2000.

[20] S. Chakrabarty and E. A. Habets, "A Bayesian approach to informed spatial filtering with robustness against doa estimation errors," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 26, no. 1, pp. 145–160, 2017.

[21] B. Liao, C. Guo, L. Huang, Q. Li, and H. C. So, "Robust adaptive beamforming with precise main beam control," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 345–356, 2017.

[22] L. Chang and C.-C. Yeh, "Performance of DMI and eigenspace-based beamformers," *IEEE Trans. Antennas Propag.*, vol. 40, no. 11, pp. 1336–1347, 1992.

[23] D. D. Feldman and L. J. Griffiths, "A projection approach for robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 867–876, 1994.

[24] A. Hassanien, S. A. Vorobyov, and K. M. Wong, "Robust adaptive beamforming using sequential quadratic programming: An iterative solution to the mismatch problem," *IEEE Signal Process. Lett.*, vol. 15, pp. 733–736, 2008.

[25] Y. Huang, M. Zhou, and S. A. Vorobyov, "New designs on MVDR robust adaptive beamforming based on optimal steering vector estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3624–3638, 2019.

[26] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, 2003.

[27] Y. Huang, S. A. Vorobyov, and Z.-Q. Luo, "Quadratic matrix inequality approach to robust adaptive beamforming for general-rank signal model," *IEEE Trans. Signal Process.*, vol. 68, pp. 2244–2255, 2020.

[28] S. A. 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, 2008.

[29] M. Rubsamen and A. B. Gershman, "Robust adaptive beamforming using multidimensional covariance fitting," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 740–753, 2011.

[30] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, 1988.

[31] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, 2008.

[32] L. Yang, M. R. McKay, and R. Couillet, "High-dimensional MVDR beamforming: Optimized solutions based on spiked random matrix models," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1933–1947, 2018.

[33] K. Elkhalil, A. Kammoun, T. Y. Al-Naffouri, and M.-S. Alouini, "Fluctuations of the SNR at the output of the MVDR with regularized Tyler estimators," *Signal Process.*, vol. 135, pp. 1–8, 2017.

[34] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, 2012.

[35] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 1–12, 2000.

[36] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, 2005.

[37] M. Zhang, A. Zhang, and Q. Yang, "Robust adaptive beamforming based on conjugate gradient algorithms," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6046–6057, 2016.

[38] A. De Maio, L. Pallotta, J. Li, and P. Stoica, "Loading factor estimation under affine constraints on the covariance eigenvalues with application to radar target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1269–1283, 2018.

[39] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3881–3885, 2012.

[40] S. Mohammadzadeh, V. H. Nascimento, R. C. de Lamare, and O. Kukrer, "Maximum entropy-based interference-plus-noise covariance matrix reconstruction for robust adaptive beamforming," *IEEE Signal Process. Lett.*, vol. 27, pp. 845–849, 2020.

[41] L. Huang, J. Zhang, X. Xu, and Z. Ye, "Robust adaptive beamforming with a novel interference-plus-noise covariance matrix reconstruction method," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1643–1650, 2015.

[42] Z. Zhang, W. Liu, W. Leng, A. Wang, and H. Shi, "Interference-plus-noise covariance matrix reconstruction via spatial power spectrum sampling for robust adaptive beamforming," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 121–125, 2015.

[43] S. Mohammadzadeh, V. H. Nascimento, R. C. de Lamare, and O. Kukrer, "Covariance matrix reconstruction based on power spectral estimation and uncertainty region for robust adaptive beamforming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 3848–3858, 2022.

[44] H. Ruan and R. C. de Lamare, "Robust adaptive beamforming using a low-complexity shrinkage-based mismatch estimation algorithm," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 60–64, 2013.

[45] ——, "Robust adaptive beamforming based on low-rank and cross-correlation techniques," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3919–3932, 2016.

[46] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[47] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, no. 9, pp. 1018–1028, 1982.

[48] S. Wang, W. Dai, and G. Y. Li, "Distributionally robust receive combining," *IEEE Trans. Signal Processing*, 2025.

[49] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[50] S. Wu and J. Zhang, "A new robust beamforming method with antennae calibration errors," in *IEEE Wireless Communications and Networking Conference*, vol. 2, 1999, pp. 869–872.

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

[52] N. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antennas Propag.*, vol. 34, no. 8, pp. 996–1012, 1986.

[53] Y. Pan, F. Duan, F. Chapeau-Blondeau, and D. Abbott, "Noise enhancement in robust estimation of location," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 1953–1966, 2018.

[54] Y. Huang, W. Yang, and S. A. Vorobyov, "Robust adaptive beamforming maximizing the worst-case SINR over distributional uncertainty sets for random INC matrix and signal steering vector," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4918–4922.

[55] K. H. Irani, Y. Huang, and S. A. Vorobyov, "SINR maximizing distributionally robust adaptive beamforming," *IEEE Trans. Signal Processing*, 2025.

[56] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, pp. 209–230, 1973.

[57] S. Ghosal and A. W. van der Vaart, *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, 2017, vol. 44.

[58] P. Stoica, R. L. Moses *et al.*, *Spectral Analysis of Signals*. Citeseer, 2005, vol. 452.

Shixiong Wang received the B.Eng. degree in detection, guidance, and control technology, and the M.Eng. degree in systems and control engineering from the School of Electronics and Information, Northwestern Polytechnical University, China, in 2016 and 2018, respectively. He received his Ph.D. degree from the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, in 2022. He is currently a Postdoctoral Research Associate with the Intelligent Transmission and Processing Laboratory, Imperial College London, London, United Kingdom, from May 2023. He was a Postdoctoral Research Fellow with the Institute of Data Science, National University of Singapore, Singapore, from March 2022 to March 2023. His research interest includes statistics and optimization theories with applications in signal processing (e.g., optimal estimation theory) and machine learning (e.g., generalization error theory).
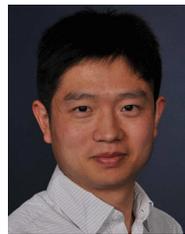
Wei Dai (Member, IEEE) received the Ph.D. degree from the University of Colorado Boulder, Boulder, Colorado, in 2007. He is currently a Senior Lecturer (Associate Professor) in the Department of Electrical and Electronic Engineering, Imperial College London, London, UK. From 2007 to 2011, he was a Postdoctoral Research Associate with the University of Illinois Urbana-Champaign, Champaign, IL, USA. His research interests include electromagnetic sensing, biomedical imaging, wireless communications, and information theory.

Geoffrey Ye Li (Fellow, IEEE) is currently a Chair Professor at Imperial College London, UK. Before joining Imperial in 2020, he was a Professor at Georgia Institute of Technology for 20 years and a Principal Technical Staff Member with AT&T Labs – Research (previous Bell Labs) for five years. He made fundamental contributions to orthogonal frequency division multiplexing (OFDM) for wireless communications, established a framework on resource cooperation in wireless networks, and introduced deep learning to communications. In these areas, he has published over 700 journal and conference papers in addition to over 40 granted patents. His publications have been cited around 80,000 times with an H-index over 130. He has been listed as a Highly Cited Researcher by Clarivate/Web of Science almost every year. Dr. Geoffrey Ye Li was elected to Fellow of the Royal Academic of Engineering (FREng), IEEE Fellow, and IET Fellow for his contributions to signal processing for wireless communications. He received 2024 IEEE Eric E. Sumner Award, 2019 IEEE ComSoc Edwin Howard Armstrong Achievement Award, and several other awards from IEEE Signal Processing, Vehicular Technology, and Communications Societies.