Robust Beamforming via Wasserstein Distance: A Data-Driven Distributionally Robust Framework

Lin Ye, Zhi Zheng, *Senior Member, IEEE*

***Abstract*—Adaptive beamforming suffers severe performance degradation from uncertainties in the array steering vector and interference-plus-noise covariance matrix. While distributionally robust optimization (DRO) provides a principled framework, existing DRO-based beamformers are often overly conservative and rely on empirically tuned parameters. This paper proposes a unified DRO beamforming framework based on the Wasserstein-2 distance. Ambiguity sets for both the random steering vector and covariance matrix are constructed as Wasserstein balls, with a physically motivated trace support constraint added to the latter to cap unrealistic worst-case power. Leveraging strong duality, the min-max problem is transformed into a tractable convex program. Without the support constraint, it reduces to an adaptive diagonal loading form whose regularization parameter is rigorously data-driven. A bootstrap-based quantile calibration strategy is developed to determine the covariance-related Wasserstein radius and support-set parameter directly from training snapshots, providing finite-sample statistical guarantees and enabling fully automatic selection of these robustness parameters. Extensive simulations demonstrate that the proposed method consistently outperforms state-of-the-art robust beamformers, establishing a new performance benchmark and systematically transitioning the design paradigm from empirical to data-driven.**

***Index Terms*—Data-driven, diagonal loading, distributionally robust optimization (DRO), robust adaptive beamforming (RAB), Wasserstein distance.**

I. Introduction

Beamforming is a cornerstone technique in array signal processing, with critical applications in radar, sonar, wireless communications, and medical imaging [1]. It works by adaptively combining sensor outputs to enhance a desired signal while suppressing interference and noise [2]–[4]. The performance benchmark is set by the minimum variance distortionless response (MVDR) beamformer, which maximizes the output signal-to-interference-plus-noise ratio (SINR) when the array steering vector and the interference-plus-noise covariance matrix are perfectly known [5]. In practice, both are fraught with uncertainty: the steering vector suffers from mismatches due to direction-of-arrival estimation errors and array imperfections, whereas at moderate to high SNR, the desired signal component dominates the sample covariance matrix, contaminating the IPN covariance estimate and leading to the well-known signal self-nulling phenomenon. This issue is further compounded when only a limited number of snapshots is available. These imperfections cause severe performance degradation, often rendering the nominal MVDR solution useless. Overcoming this limitation is the central goal of robust adaptive beamforming (RAB) research over the past four decades [6].

Classical robust approaches, such as diagonal loading [7] and worst-case performance optimization [8], [9], mitigate steering vector mismatches by modeling them as deterministic uncertainty sets—typically norm balls or ellipsoids [10]—and ensure a distortionless response for all vectors within these sets. Despite their practical success, these methods provide no systematic framework for handling the uncertainty inherent in the IPN covariance matrix. As a result, their performance inevitably saturates under high signal-to-noise ratio (SNR) conditions, where the desired signal contamination of the sample covariance matrix becomes the dominant source of performance loss. More fundamentally, the robustness parameters in these methods, such as the diagonal loading factor or the size of the uncertainty set, are typically chosen empirically [11], lacking a principled mechanism to adapt to the actual level of uncertainty present in the observed data.

Distributionally robust optimization (DRO) has recently emerged as a principled framework that treats uncertainties probabilistically by constructing ambiguity sets for the unknown distributions of the steering vector and the IPN covariance matrix [12]–[14]. In particular, recent DRO-based methods can jointly model dual uncertainties through moment constraints [15]. However, two fundamental limitations persist. First, moment-based ambiguity sets, by definition, contain all distributions whose first- and second-order moments equal prescribed values. This set inevitably includes many distributions that are physically implausible, resulting in excessive conservatism. Second, the parameters governing these ambiguity sets are still chosen empirically, without rigorous finite-sample statistical justification. These limitations motivate a more precise, physically meaningful, and data-adaptive framework.

This paper proposes a unified distributionally robust beamforming framework based on the Wasserstein-2 distance. The ambiguity sets for both the random steering vector and the random IPN covariance matrix are constructed as Wasserstein balls, with the covariance set further equipped by a physically motivated trace support constraint. Leveraging strong duality theory [16], the original min-max problem is transformed into tractable convex programs. Without the support constraint, the

Lin Ye and Zhi Zheng are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 202422011014@std.uestc.edu.cn; zz@uestc.edu.cn).

problem reduces to an adaptive diagonal loading form, revealing that the optimal regularization parameter is exactly the data-driven Wasserstein radius. With the support constraint active, the problem becomes a mixed semidefinite–second-order cone program, solved via a difference-of-convex-functions algorithm for rank-one recovery. Building on the foundational finite-sample theory for Wasserstein distributionally robust optimization established in [16], [17], a bootstrap-based quantile calibration strategy is developed to automatically determine the Wasserstein radius and the support-set parameter for the covariance matrix directly from training snapshots, providing rigorous finite-sample statistical guarantees under a prescribed confidence level. For the steering vector uncertainty, the corresponding Wasserstein radius can be conservatively estimated from physical error bounds of the array system, thereby ensuring worst-case robustness [18]. Together, these components enable fully automatic, physically grounded selection of robustness parameters, transitioning the design of robust beamformers from an empirical, trial-and-error process to a principled, data-driven paradigm.

In parallel with this work, several recent contributions have significantly advanced the understanding and application of DRO for robust beamforming. A unified DRO framework [19] demonstrated that many existing robust methods, such as diagonal loading, eigenvalue thresholding, and prior-knowledge embedding, can be interpreted as special cases of distributionally robust optimization. However, this framework primarily provides a theoretical unification rather than a practical solution to the problem of optimally selecting the regularization parameters—the critical question of how much loading or thresholding should be applied remains unresolved. Another recent paper [15] addressed the dual uncertainties in both the steering vector and the covariance matrix through moment constraints. While this method jointly models both sources of uncertainty, the moment-based ambiguity sets inherently accommodate all distributions matching prescribed moments, resulting in excessive conservatism that degrades beamformer performance. Moreover, the performance of this method is highly sensitive to the quality of prior information—simulations in [15] construct the angle sector for moment estimation centered at the true direction-of-arrival, an idealized condition rarely met in practice. When the sector is centered at the presumed (mismatched) direction, the performance degrades substantially. The Wasserstein distance has recently been explored for RAB in two related efforts: the work in [20] that employed the Wasserstein-1 distance to separately model the steering vector and INC matrix uncertainties, and the work in [21] that unrolled a FISTA-based solver into a trainable network and incorporated a neural module for INC matrix reconstruction. The present framework differs from all the above works in several fundamental respects. First, rather than unifying existing methods or handling uncertainties separately, we formulate a unified minimax problem that simultaneously accounts for both steering vector and covariance matrix distributional uncertainty under the Wasserstein-2 metric. Second, we introduce a physically motivated trace support constraint on the covariance ambiguity set, which prevents the worst-case INC matrix from exceeding physically realizable power levels, thereby overcoming the excessive conservatism that plagues moment-based approaches. Most importantly, the proposed bootstrap-based quantile calibration strategy provides rigorous finite-sample statistical guarantees for the covariance-related parameters, directly answering the open question of optimal regularization parameter selection that the aforementioned unified framework leaves unresolved. Together, these components transition the design of robust beamformers from an empirical, trial-and-error process to a principled, data-driven paradigm.

The remainder of this paper is structured as follows. Section II introduces the system model and formulates the distributionally robust beamforming problem under the Wasserstein ambiguity sets. Section III leverages strong duality to transform the intractable min-max problem into tractable convex programs, distinguishing between the cases with and without the support constraint. Section IV details the proposed bootstrap-based calibration strategy that automatically determines the Wasserstein radii with finite-sample statistical guarantees, directly addressing the core issue of parameter selection. Section V provides comprehensive simulations to validate the superiority of the proposed method against state-of-the-art alternatives under various mismatch scenarios. Finally, Section VI concludes the paper.

Notations: To ensure a clear and unambiguous presentation in the following sections, we summarize the key notations used throughout this paper. Scalars are denoted by italic letters, vectors by boldface italic lowercase letters, matrices by boldface italic uppercase letters, and sets or spaces by calligraphic or blackboard bold letters. Mathematical operators and constants are typeset in upright (non-italic) font. Specifically, $\mathcal{H}^N$ denotes the set of $N \times N$ Hermitian matrices, and $\mathcal{M}$ the set of all probability measures on a given measurable space. The transpose, conjugate transpose, and complex conjugate are denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. The trace and rank of a matrix are $\mathrm{tr}(\cdot)$ and $\mathrm{rank}(\cdot)$. We use $\mathrm{vec}(\cdot)$ for the standard vectorization operator and $\mathrm{vec}_{\mathbb{R}}(\cdot)$ for its real-valued counterpart. The Euclidean and Frobenius norms are $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. For a Hermitian matrix, $\lambda_{\max}(\cdot)$ returns its maximum eigenvalue, and $[\boldsymbol{A}]_+$ is the projection of $\boldsymbol{A}$ onto the positive semidefinite cone. Mathematical expectation is denoted by $\mathbb{E}[\cdot]$, and $\mathrm{Quantile}_{1-\alpha}(\cdot)$ represents the $(1-\alpha)$-th empirical quantile of a sample set.

II. Signal Model and Problem Formulation

We consider a narrowband signal scenario where an antenna array composed of N elements receives a far-field source signal $s(t)$. The signal impinges on the array from a specific direction, characterized by the complex-valued steering vector $\boldsymbol{a} \in \mathbb{C}^N$. Consequently, the spatial signature of the signal of interest (SOI) at the array is given by:

$$\boldsymbol{s}(t) = s(t)\boldsymbol{a}. \tag{1}$$

The measured snapshot vector $\boldsymbol{y}(t) \in \mathbb{C}^N$ at the array output is a superposition of the SOI, potential interfering signals, and background sensor noise. This relationship is expressed as:

$$\boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{i}(t) + \boldsymbol{n}(t), \tag{2}$$

where $\boldsymbol{i}(t)$ and $\boldsymbol{n}(t)$ denote the interference component and additive noise vector, respectively. These components are assumed to be statistically independent from each other and from the SOI.

The core operation of beamforming is to linearly combine the signals from all antenna elements using a complex weight vector $\boldsymbol{w} \in \mathbb{C}^N$. The final array output $x(t)$ is thus obtained by:

$$x(t) = \boldsymbol{w}^H \mathbf{y}(t). \tag{3}$$

Based on the observed snapshot vector and the expression of the desired signal, the output signal can be further expressed as:

$$x(t) = s(t)\boldsymbol{w}^H \boldsymbol{a} + \boldsymbol{w}^H \big(\boldsymbol{i}(t) + \boldsymbol{n}(t)\big). \tag{4}$$

Therefore, the output SINR is given by:

$$\mathrm{SINR} = \frac{\sigma_{\mathrm{s}}^2 |\boldsymbol{w}^H \boldsymbol{a}|^2}{\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}}, \tag{5}$$

where σ_s^2 is the power of the desired signal, and $\boldsymbol{R}_{i+n} = \mathbb{E}\{(\boldsymbol{i}(t) + \boldsymbol{n}(t))(\boldsymbol{i}(t) + \boldsymbol{n}(t))^H\} \in \mathbb{C}^{N \times N}$ denotes the interference-plus-noise covariance matrix. In order to maximize the SINR, we formulate the following optimization problem:

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \boldsymbol{R}_{i+n} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^H \boldsymbol{a} = 1. \tag{6}$$

However, in practical scenarios, two primary challenges arise. On one hand, since the training snapshots are contaminated by the desired signal, we cannot obtain a pure interference-plus-noise covariance matrix $\boldsymbol{R}_{i+n}$. On the other hand, factors such as direction-of-arrival (DOA) estimation errors and array position errors cause a mismatch between the presumed steering vector and the true one. Consequently, a distributionally robust adaptive beamforming optimization problem based on the Wasserstein distance can be formulated as follows:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \max_{G_1 \in \mathcal{D}_1^w} \mathbb{E}_{G_1}\left[\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}\right] \\ \text{s.t.} \quad & \min_{G_2 \in \mathcal{D}_2^w} \mathbb{E}_{G_2}\left[|\boldsymbol{w}^H \boldsymbol{a}|^2\right] \geq \alpha. \end{aligned} \tag{7}$$

where $\mathcal{D}_1^w$ and $\mathcal{D}_2^w$ denote the distributional ambiguity sets for the interference-plus-noise covariance matrix and the steering vector, respectively, both constructed based on the Wasserstein distance. Specifically, the set $\mathcal{D}_1^w$ is defined as follows:

$$\mathcal{D}_1^w = \left\{ G_1 \in \mathcal{M}_1 \middle| \begin{array}{l} \mathbb{P}_{G_1}(\boldsymbol{R} \in \mathcal{Z}_1) = 1, \\ W_2^2(G_1, \hat{G}_1) \leq \rho_1 \end{array} \right\}, \tag{8}$$

where $\mathcal{M}_1$ denotes the set of all probability measures on the measurable space $(\mathcal{H}^N, \mathcal{B}_1)$, $\mathcal{B}_1$ is the Borel σ-algebra on $\mathcal{H}^N$, and $\mathcal{Z}_1 \subseteq \mathcal{H}^N$ is a Borel set specifying the support of the distribution. The nominal distribution $\hat{G}_1$ is typically chosen as the empirical measure derived from available training data, and $\rho_1 > 0$ is the squared Wasserstein distance that quantifies the degree of distributional uncertainty.

The distributional set $\mathcal{D}_1^w$ is constructed within the distributionally robust optimization framework based on the Wasserstein distance. The first constraint ensures that all considered distributions are supported on the prescribed Borel set $\mathcal{Z}_1$, thereby incorporating prior structural knowledge and physical limitations, which guarantees practical physical significance (e.g., trace constraints). The second constraint restricts the ambiguity set to distributions whose squared Wasserstein distance from the nominal distribution $\hat{G}_1$ does not exceed ρ_1, effectively controlling the size of the uncertainty set and ensuring robustness against statistical estimation errors. By combining the support condition with the Wasserstein-ball constraint, $\mathcal{D}_1^w$ captures a broad yet well-behaved family of distributions, enabling tractable robust optimization while maintaining key probabilistic guarantees.

Although the true distribution of the steering vector is unknown, it is assumed to reside within a Wasserstein ball centered at the nominal distribution $P_0 = \delta_{\boldsymbol{a}_0}$ (i.e., $\boldsymbol{a}$ Dirac delta distribution), where $\boldsymbol{a}_0$ is the presumed steering vector. Accordingly, the uncertainty set for the steering vector, denoted as $\mathcal{D}_2^w$, can be constructed in a similar manner as follows:

$$\mathcal{D}_2^w = \left\{ G_2 \in \mathcal{M}_2 \middle| \begin{array}{l} \mathbb{P}_{G_2}(\boldsymbol{a} \in \mathcal{Z}_2) = 1, \\ W_2^2(G_2, \boldsymbol{a}_0) \leq \epsilon_a \end{array} \right\}, \tag{9}$$

where $\mathcal{M}_2$ defines the Wasserstein uncertainty set $\mathcal{D}_2^w$ for the steering vector. Here, $\mathcal{M}_2$ denotes the set of all probability measures on the measurable space $(\mathbb{C}^N, \mathcal{B}_2)$, where $\mathcal{B}_2$ is the Borel σ-algebra on $\mathbb{C}^N$, and $\mathcal{Z}_2 \subseteq \mathbb{C}^N$ defines the support of admissible steering vectors. The nominal distribution is the Dirac measure $\delta_{\boldsymbol{a}_0}$ concentrated at the presumed steering vector $\boldsymbol{a}_0$, and the parameter $\epsilon_a > 0$ is the upper bound on the squared Wasserstein distance, thereby quantifying the permissible degree of distributional deviation.

The uncertainty set $\mathcal{D}_2^w$ provides a probabilistic description for steering vector mismatch. Its construction rests on two pillars: the support condition and the distributional proximity condition. The support condition $\mathbb{P}_{G_2}(\boldsymbol{a} \in \mathcal{Z}_2) = 1$ ensures the physical realizability of $\boldsymbol{a}$ (e.g., satisfying norm bounds or sector constraints) by restricting its realizations to the predefined set $\mathcal{Z}_2$. The proximity condition $W_2^2(G_2, \boldsymbol{a}_0) \leq \epsilon_a$ utilizes the Wasserstein metric to confine the true distribution of $\boldsymbol{a}$ within a specified distance from the nominal point-mass distribution. This dual-condition structure ensures robustness against a well-defined family of distributional perturbations. Consequently, $\mathcal{D}_2^w$ effectively captures uncertainties arising from array calibration errors, inaccuracies in direction-of-arrival estimation, and other practical imperfections in the array manifold.

In distributionally robust beamforming, selecting an appropriate probabilistic distance to construct the distributional ambiguity set is of paramount importance. This paper adopts the Wasserstein distance (also known as the earth mover's distance) as the measure for distributional uncertainty. Its core advantage lies in quantifying the minimum "work" required to transport one probability distribution to another, thereby enabling the capture of differences in the overall shape between two distributions, rather than merely deviations in moment information such as the mean or covariance. Under appropriate support set constraints, the ambiguity set based on the Wasserstein distance, via its duality theory, often allows derivation of tractable convex optimization problems, offering a mathematical framework that achieves a good balance between model fidelity and computational tractability. Simultaneously, its geometric intuitiveness facilitates flexible adjustment of the robustness strength through the radius parameters.

III. Dual Reformulation and Convex Optimization

A. Dual Reformulation for the Objective (Covariance Matrix)

In this section, we derive the dual reformulation of the distributionally robust beamforming problem presented in (7). The core of the derivation involves replacing the inner maximization and minimization problems in the objective and the constraint, respectively, with their corresponding dual forms based on the strong duality of Wasserstein distributionally robust optimization. This transformation converts the original semi-infinite, min-max problem into a tractable convex optimization problem, paving the way for efficient numerical solution.

We begin by addressing the inner maximization problem in the objective of problem (7), where the uncertainty set $\mathcal{D}_1^w$ governs the distributions of the interference-plus-noise covariance matrix $\boldsymbol{R}$ (for notational brevity in the following derivation, we denote $\boldsymbol{R} = \boldsymbol{R}_{i+n}$). First, we present the definition of the Wasserstein distance. Here, we specifically choose the Wasserstein-2 distance, which is defined as follows:

$$W_2(\mu,\nu) = \left(\inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\,\mathrm{d}\gamma(x,y) \right)^{1/2}, \quad (10)$$

where μ and ν are two probability distributions defined on a space $\mathcal{X}$, $\Pi(\mu,\nu)$ denotes the set of all joint distributions (couplings) γ with marginals μ and ν, and $c(x,y) : \mathcal{X}\times\mathcal{X} \to [0,\infty)$ is a cost function. For the second-order distance W_2, the cost function is typically chosen as the squared Euclidean norm, i.e., $c(x,y) = \|x-y\|^2$. where $c(x,y)$ is the cost function. For matrices x and y, we define it as the squared Frobenius norm of their difference, i.e., $c(x,y) = \|x-y\|_F^2$. Therefore, the inner maximization problem in the original objective function of problem (7) can be reformulated as:

$$\sup_{G_1,\gamma} \left\{ \int_{\mathcal{Z}_1\times\mathcal{Z}_1} f(\boldsymbol{R}')\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) \;\middle|\; \begin{array}{l} \gamma\in\Pi(G_1,\widehat{G}_1) \\ G_1\in\mathbb{P}(\mathcal{Z}_1) \\ \int_{\mathcal{Z}_1\times\mathcal{Z}_1} \|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) \le \rho_1 \end{array} \right\} \quad (11)$$

where $f(\boldsymbol{R}) = \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}$ and γ denotes the joint probability density distribution, and $\mathbb{P}(\mathcal{Z}_1)$ denotes the set of probability distributions on the support set Z_1.Therefore, the Lagrangian dual function of problem (11) is expressed as follows:

$$L(G_1,\gamma;\lambda) = \int_{\mathcal{Z}_1\times\mathcal{Z}_1} f(\boldsymbol{R}')\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) - \lambda\left(\int_{\mathcal{Z}_1\times\mathcal{Z}_1} \|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) - \rho_1 \right), \quad (12)$$

where the Lagrange multiplier $\lambda\in\mathbb{R}$ and $\lambda\ge 0$. Rearranging the terms, the Lagrangian can be rewritten as follows:

$$L(\gamma;\lambda) = \int_{\mathcal{Z}_1\times\mathcal{Z}_1} f(\boldsymbol{R}')\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) - \lambda\left(\int_{\mathcal{Z}_1\times\mathcal{Z}_1} \|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\,\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) - \rho_1 \right). \quad (13)$$

Here, G_1 is omitted because it is implicitly contained in γ (the marginal distribution of γ determines G_1). Therefore, based on the relationship between the integral value of the cost function and the parameter ρ_1, the following piecewise function can be obtained:

$$\inf_{\lambda\ge 0} L(\gamma;\lambda) = \begin{cases} \mathbb{E}_\gamma[f(\boldsymbol{R}')] & \text{if } \mathbb{E}_\gamma[\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2] \le \rho_1 \\ -\infty & \text{if } \mathbb{E}_\gamma[\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2] > \rho_1 \end{cases}, \quad (14)$$

then the original problem (i.e., the inner maximization problem of the objective function) is equivalent to:

$$\sup_{\gamma\in\Pi(G,\widehat{G}_1),\;\; G_1\in\mathbb{P}(\mathcal{Z}_1)} \left[\inf_{\lambda\ge 0} L(\gamma;\lambda) \right]. \quad (15)$$

It follows from the definition that:

$$\sup_{\substack{\gamma\in\Pi(G_1,\widehat{G}_1),\\ G_1\in\mathbb{P}(\mathcal{Z}_1)}} \left[\inf_{\lambda\ge 0} L(\gamma;\lambda) \right] \le \inf_{\lambda\ge 0} \left[\sup_{\substack{\gamma\in\Pi(G_1,\widehat{G}_1),\\ G_1\in\mathbb{P}(\mathcal{Z}_1)}} L(\gamma;\lambda) \right]. \quad (16)$$

Since the original problem is a convex optimization problem and its linear inequality constraints admit a strictly feasible solution, the Slater condition is satisfied. Consequently, strong duality holds, and the duality gap is zero. Thus, the original problem is equivalent to:

$$\inf_{\lambda\ge 0} \left\{ \sup_{\substack{\gamma\in\Pi(G_1,\widehat{G}_1),\\ G_1\in\mathbb{P}(\mathcal{Z}_1)}} \left[\lambda\rho_1 + \int_{\mathcal{Z}_1\times\mathcal{Z}_1} \left(f(\boldsymbol{R}') - \lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2 \right) \mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) \right] \right\}. \quad (17)$$

Since the inner maximization problem is independent of the term $\lambda\rho_1$ (which can be viewed as a constant), it is equivalent to:

$$\sup_{\substack{\gamma\in\Pi(G_1,\widehat{G}_1),\\ G_1\in\mathbb{P}(\mathcal{Z}_1)}} \int_{\mathcal{Z}_1\times\mathcal{Z}_1} \left(f(\boldsymbol{R}') - \lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2 \right) \mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}). \quad (18)$$

Notice that $\mathrm{d}\gamma(\boldsymbol{R}',\boldsymbol{R}) = \mathrm{d}\gamma(\boldsymbol{R}' \mid \boldsymbol{R})\,\mathrm{d}\widehat{G}_1(\boldsymbol{R})$, and the marginal distribution $\widehat{G}_1$ is fixed, we can first fix $\boldsymbol{R}$ in $\widehat{G}_1$ and then choose the conditional distribution $\mathrm{d}\gamma(\boldsymbol{R}' \mid \boldsymbol{R})$ for each $\boldsymbol{R}$ to maximize the expectation of the joint distribution. In fact, the optimal $\mathrm{d}\gamma(\boldsymbol{R}' \mid \boldsymbol{R})$ will concentrate all its mass on the $\boldsymbol{R}'$ that maximizes $f(\boldsymbol{R}') - \lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2$ (for a fixed $\boldsymbol{R}$). In other words, according to optimal transport theory, the inner maximization problem is equivalent to:

$$\int_{\mathcal{Z}_1} \sup_{\boldsymbol{R}'\in\mathcal{Z}_1} \left(f(\boldsymbol{R}') - \lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2 \right) \mathrm{d}\widehat{G}_1(\boldsymbol{R}), \quad (19)$$

this integral expression with respect to the nominal distribution

$\widehat{G}_1(\boldsymbol{R})$ can be rewritten as follows:

$$\mathbb{E}_{\widehat{G}_1}\left[\sup_{\boldsymbol{R}'\in\mathcal{Z}_1}\left(f(\boldsymbol{R}')-\lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\right)\right]. \tag{20}$$

Then the inner maximization problem of the objective function in problem (7) reduces to:

$$\inf_{\lambda\geq 0}\left\{\lambda\rho_1+\mathbb{E}_{\widehat{G}_1}\left[\sup_{\boldsymbol{R}'\in\mathcal{Z}_1}\left(f(\boldsymbol{R}')-\lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\right)\right]\right\}. \tag{21}$$

To advance the solution, the following analysis will separately consider two cases: without support set constraints (i.e., $\mathcal{Z}_1=\mathbb{H}^N$) and with support set constraints, for the inner supremum problem.

Case 1: Ignoring the Support Set (Unconstrained Approximation)

When the support set $\mathcal{Z}_1$ is ignored, the inner maximization problem in problem (21) reduces to:

$$\sup_{\boldsymbol{R}'\in\mathbb{H}^N}\left(\boldsymbol{w}^H\boldsymbol{R}'\boldsymbol{w}-\lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2\right). \tag{22}$$

Let $g(\boldsymbol{R}')=\boldsymbol{w}^H\boldsymbol{R}'\boldsymbol{w}-\lambda\|\boldsymbol{R}'-\boldsymbol{R}\|_F^2$, and taking its derivative yields:

$$\nabla_{\boldsymbol{R}'}g(\boldsymbol{R}')=\boldsymbol{w}\boldsymbol{w}^H-2\lambda(\boldsymbol{R}'-\boldsymbol{R}). \tag{23}$$

Setting the derivative to zero, we thus obtain the optimal solution:

$$\boldsymbol{R}'^{*}=\boldsymbol{R}+\frac{1}{2\lambda}\boldsymbol{w}\boldsymbol{w}^H, \tag{24}$$

this closed-form solution indicates that, when the support set constraint is ignored, the optimal solution $\boldsymbol{R}'^{*}$ of the inner maximization problem is the original sample covariance matrix $\boldsymbol{R}$ plus a rank-one correction term $\frac{1}{2\lambda}\boldsymbol{w}\boldsymbol{w}^H$ formed by the beamforming vector $\boldsymbol{w}$, where the regularization parameter λ controls the strength of the correction. Therefore, substitution shows that problem (22) is equivalent to:

$$\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}+\frac{1}{4\lambda}\|\boldsymbol{w}\|^4, \tag{25}$$

where $\|\boldsymbol{w}\|^4=(\|\boldsymbol{w}\|_2^2)^2$.Suppose the nominal distribution $\widehat{G}_1$ satisfies that its expectation is $\boldsymbol{S}_0$, where $\boldsymbol{S}_0$ can be the sample covariance matrix $\boldsymbol{R}=\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{y}(k)\boldsymbol{y}^H(k)$ with K snapshots, or, more precisely, the reconstructed interference-plus-noise covariance matrix $\boldsymbol{R}_{i+n}$.Therefore, in the absence of support set constraints, problem (21) is equivalent to:

$$\inf_{\lambda\geq 0}\left(\lambda\rho_1+\boldsymbol{w}^H\boldsymbol{S}_0\boldsymbol{w}+\frac{1}{4\lambda}\|\boldsymbol{w}\|^4\right), \tag{26}$$

let $h(\lambda)=\lambda\rho_1+\boldsymbol{w}^H\boldsymbol{S}_0\boldsymbol{w}+\frac{1}{4\lambda}\|\boldsymbol{w}\|^4$, and taking its derivative yields:

$$\nabla_\lambda h(\lambda)=\rho_1-\frac{1}{4\lambda^2}\|\boldsymbol{w}\|^4, \tag{27}$$

setting the derivative to zero yields the optimal solution:

$$\lambda^{*}=\frac{\|\boldsymbol{w}\|_2^2}{2\sqrt{\rho_1}}, \tag{28}$$

substituting this result yields that problem (26) reduces to:

$$\boldsymbol{w}^H\boldsymbol{S}_0\boldsymbol{w}+\sqrt{\rho_1}\,\|\boldsymbol{w}\|_2^2. \tag{29}$$

Therefore, ignoring the support set constraints, the objective function of optimization problem (7) ultimately reduces to:

$$\min_{\boldsymbol{w}}\ \boldsymbol{w}^H\boldsymbol{S}_0\boldsymbol{w}+\sqrt{\rho_1}\,\|\boldsymbol{w}\|_2^2. \tag{30}$$

Discussion

The reformulation of the original distributionally robust problem into the compact form (30) provides a key theoretical insight: the requirement for robustness against distributional errors in the channel is exactly equivalent to penalizing the norm of the beamformer. This reveals the intrinsic mechanism of the Wasserstein DRO framework—it automatically enforces a trade-off between minimizing the nominal output power and reducing the beamformer's sensitivity to steering vector mismatches.

The fundamental distinction from traditional diagonal loading [7], [11] lies in the explicit physical and statistical basis of our final expression. In conventional robust adaptive beamforming, the diagonal loading factor is typically treated as a heuristic hyper-parameter tuned empirically for performance. Although diagonal loading is recognized to act as a regularizer that enhances robustness by effectively reducing the weight vector norm (or improving the condition number of the covariance matrix), the underlying mechanism is that the sensitivity of the array response $\boldsymbol{w}^H\boldsymbol{a}$ to perturbations in the steering vector $\boldsymbol{a}$ is governed by $\|\boldsymbol{w}\|$, since $\nabla_{\boldsymbol{a}}(\boldsymbol{w}^H\boldsymbol{a})=\boldsymbol{w}$. Thus, penalizing $\|\boldsymbol{w}\|$ directly reduces this sensitivity. However, the critical flaw of the traditional approach is the lack of a principled method to determine the optimal amount of loading, making it difficult to systematically balance robustness (conservatism) against performance.

In contrast, the regularization parameter in our formulation, $\sqrt{\rho_1}$, is not an arbitrary hyper-parameter. It is derived directly from the Wasserstein uncertainty set, where ρ_1 is the prescribed upper bound on the squared Wasserstein distance. Thus, the robustness level is intrinsically tied to the statistical uncertainty quantified by the Wasserstein framework. This radius has a precise statistical interpretation: it quantifies the size of the distributional uncertainty set. As established in the foundational work of Blanchet et al [17], ρ_1 can be rigorously calibrated from data (e.g., via a quantile-based method) to guarantee with high probability that the true channel distribution lies within the uncertainty set. The specific calibration method for ρ_1 will be detailed in the section IV. Therefore, our framework not only elucidates why regularization (diagonal loading) works but also provides a rigorous theory to determine its theoretically optimal strength. This transitions the design of robust beamformers from an empirical, trial-and-error process to a principled, theoretically-automated, and data-driven procedure.

Case 2: Considering the Support Set (Constrained Optimization)

In the above analysis, the constraint on the support set was ignored. We now investigate whether the constraints on the support set are satisfied by the obtained optimal solution $\boldsymbol{R}'^{*}$ defined in (24).

Consider the support set defined as $\mathcal{Z}_1 = \{\boldsymbol{R} \in \mathbb{H}^N \mid \boldsymbol{R} \succeq 0, \operatorname{tr}(\boldsymbol{R}) \leq \rho_2\}$. It is clear that the first constraint (positive semidefiniteness) is satisfied. However, the second constraint, namely, the trace (or energy) constraint $\operatorname{tr}(\boldsymbol{R}) \leq \rho_2$, may be violated. Specifically, $\operatorname{tr}(\boldsymbol{R}'^*) = \operatorname{tr}(\boldsymbol{R}) + \frac{1}{2\lambda}\|\boldsymbol{w}\|^2$. The term $\frac{1}{2\lambda}\|\boldsymbol{w}\|^2$ becomes unbounded as $\lambda \to 0^+$, which causes the trace of $\boldsymbol{R}'^*$ to diverge. Even for a small but finite λ, since $\|\boldsymbol{w}\|^2 > 0$, the trace $\operatorname{tr}(\boldsymbol{R}'^*)$ can significantly exceed $\operatorname{tr}(\boldsymbol{R})$, potentially resulting in $\operatorname{tr}(\boldsymbol{R}'^*) > \rho_2$.

This violation reveals a fundamental limitation of conventional diagonal loading. Although adding a uniform loading factor δ (i.e., $\boldsymbol{R}' = \boldsymbol{R} + \delta\mathbf{I}$) can enhance robustness, it indiscriminately increases the total power $\operatorname{tr}(\boldsymbol{R}')$. When δ is set too large, the inner maximization problem considers an unrealistic interference scenario with excessive power, far beyond the physical limit ρ_2. Consequently, the outer beamformer design becomes overly conservative, leading to performance degradation in practical, power-limited environments. This explains why traditional methods struggle to balance robustness and performance—they lack a mechanism to optimally tie the loading factor to the physical power constraint.

Violating the constraint on the support set therefore leads to the following critical drawbacks:

1) **Physical unrealizability:** The solution corresponds to an interference-plus-noise power that exceeds the system's feasible upper bound, contradictating physical realizability.
2) **Overly conservative design:** Using such an unrealistic high-power scenario in the inner maximization forces the outer beamformer to be designed for an excessively pessimistic case, resulting in unnecessary performance loss.

In summary, the original intention of distributionally robust optimization is to find the worst-case scenario within a reasonable uncertainty set. The support set constraint defines the physical boundary of this set. Ignoring it implies allowing unreasonable extreme cases, thereby rendering the optimization problem practically meaningless. The formulation in Case 2, which explicitly incorporates the support set $\mathcal{Z}_1$, ensures that the robust beamformer is derived from a physically admissible worst-case scenario, thereby overcoming the inherent trade-off dilemma of traditional diagonal loading.

Next, we derive the transformation of the original problem when the support set is considered. Analogously, for each fixed $\boldsymbol{R}$ drawn from the nominal distribution $\widehat{G}_1$, we independently maximize its conditional expectation. We still search for the optimal $\boldsymbol{R}'^*$ within G_1, but now $\boldsymbol{R}'^*$ must belong to the support set $\mathcal{Z}_1$. Therefore, for each fixed R, the inner maximization in problem (21) reduces to:

$$\sup_{\boldsymbol{R}' \in \mathcal{Z}_1} \left(\boldsymbol{w}^H \boldsymbol{R}' \boldsymbol{w} - \lambda \|\boldsymbol{R}' - \boldsymbol{R}\|_F^2\right). \tag{31}$$

This is a convex optimization problem with the convex constraint Z_1. For the nominal distribution $\widehat{G}_1$ (composed of K samples $\boldsymbol{R}_k$, as detailed in the section IV), problem (31) transforms from taking the expectation over a continuous distribution to the average over a finite set of samples:

$$\frac{1}{K}\sum_{k=1}^{K} \sup_{\boldsymbol{R}' \in \mathcal{Z}_1} \left(\boldsymbol{w}^H \boldsymbol{R}' \boldsymbol{w} - \lambda \left\|\boldsymbol{R}' - \boldsymbol{R}_k\right\|_F^2\right), \tag{32}$$

therefore, problem (21) transforms into the following form:

$$\inf_{\lambda \geq 0} \left\{ \lambda\rho_1 + \frac{1}{K}\sum_{k=1}^{K} \sup_{\boldsymbol{R}' \in \mathcal{Z}_1} \left(\boldsymbol{w}^H \boldsymbol{R}' \boldsymbol{w} - \lambda\|\boldsymbol{R}' - \boldsymbol{R}_k\|_F^2\right) \right\}, \tag{33}$$

for each sample $\boldsymbol{R}_k$, the following problem needs to be solved:

$$\sup_{\boldsymbol{R}' \in \mathcal{Z}_1} \left(\boldsymbol{w}^H \boldsymbol{R}' \boldsymbol{w} - \lambda \left\|\boldsymbol{R}' - \boldsymbol{R}_k\right\|_F^2\right), \tag{34}$$

expanding the Frobenius norm yields:

$$\begin{aligned} \left\|\boldsymbol{R}' - \boldsymbol{R}_k\right\|_F^2 &= \operatorname{tr}\Big(\left(\boldsymbol{R}' - \boldsymbol{R}_k\right)^H \left(\boldsymbol{R}' - \boldsymbol{R}_k\right)\Big) \\ &= \operatorname{tr}\big(\boldsymbol{R}'^2\big) - 2\operatorname{tr}\big(\boldsymbol{R}'\boldsymbol{R}_k\big) + \operatorname{tr}\big(\boldsymbol{R}_k^2\big), \end{aligned} \tag{35}$$

furthermore, since $\boldsymbol{w}^H \boldsymbol{R}' \boldsymbol{w} = \operatorname{tr}(\boldsymbol{R}' \boldsymbol{w}\boldsymbol{w}^H)$, the problem (34) can be simplified as follows:

$$\sup_{\boldsymbol{R}' \in \mathcal{Z}_1} \left\{ \operatorname{tr}\left(\boldsymbol{R}'(\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k)\right) - \lambda\operatorname{tr}(\boldsymbol{R}'^2) \right\} - \lambda\operatorname{tr}(\boldsymbol{R}_k^2). \tag{36}$$

For the support set $\mathcal{Z}_1 = \{\boldsymbol{R} \in \mathbb{H}^N \mid \boldsymbol{R} \succeq 0, \operatorname{tr}(\boldsymbol{R}) \leq \rho_2\}$, we introduce a Lagrange multiplier $\mu \geq 0$ to handle the trace constraint. Then, following a similar dual reconstruction derivation as before, after simplification and rearrangement, problem (36) is equivalent to:

$$\begin{aligned} \sup_{\boldsymbol{R}' \succeq 0} \Bigg\{ \inf_{\mu \geq 0} \Big(& \operatorname{tr}\big(\boldsymbol{R}'(\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k)\big) \\ & -\lambda\operatorname{tr}(\boldsymbol{R}'^2) - \mu\operatorname{tr}(\boldsymbol{R}') \\ & +\mu\rho_2 - \lambda\operatorname{tr}(\boldsymbol{R}_k^2)\Big)\Bigg\}, \end{aligned} \tag{37}$$

since (36) is a convex maximization problem (the objective is strictly concave in $\boldsymbol{R}'$ and the constraint is linear) and Slater's condition holds (e.g., $\boldsymbol{R}' = \frac{\rho_2}{2N}\boldsymbol{I}$ is strictly feasible with $\boldsymbol{R}' \succ 0$ and $\operatorname{tr}(\boldsymbol{R}') < \rho_2$), strong duality is guaranteed [22]. Consequently, interchanging the order of the infimum and supremum is justified [23], and problem (36) is equivalent to:

$$\begin{aligned} \inf_{\mu \geq 0} \Bigg\{ \mu\rho_2 + \sup_{\boldsymbol{R}' \succeq 0} \Big(& \operatorname{tr}\Big(\boldsymbol{R}'(\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I})\Big) \\ & - \lambda\operatorname{tr}(\boldsymbol{R}'^2)\Big)\Bigg\} - \lambda\operatorname{tr}(\boldsymbol{R}_k^2), \end{aligned} \tag{38}$$

the inner maximization problem of this formulation reduces, for a fixed μ, to solving the following problem:

$$\sup_{\boldsymbol{R}' \succeq 0} \Big(\operatorname{tr}\big(\boldsymbol{R}'\boldsymbol{B}_k(\mu)\big) - \lambda\operatorname{tr}\big(\boldsymbol{R}'^2\big)\Big), \tag{39}$$

where $\boldsymbol{B}_k = \boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}$ (for notational brevity in the following derivation, we denote $\boldsymbol{B}_k = \boldsymbol{B}_k(\mu)$). This is a strictly concave quadratic function in $\boldsymbol{R}'$ defined over the positive semidefinite cone. Therefore, the optimal solution to problem (39) is given by (See Appendix A):

$$\boldsymbol{R}'^{*} = \frac{1}{2\lambda}[\boldsymbol{B}_k]_+, \quad (40)$$

where $[\boldsymbol{X}]_+$ denotes the projection of the matrix $\boldsymbol{X}$ onto the positive semidefinite cone.Substituting this optimal solution into problem (39) yields that it is equivalent to:

$$\frac{1}{2\lambda}\operatorname{tr}\left([\boldsymbol{B}_k]_+\boldsymbol{B}_k\right) - \frac{1}{4\lambda}\operatorname{tr}\left([\boldsymbol{B}_k]_+^2\right), \quad (41)$$

since $[\boldsymbol{B}_k]_+\boldsymbol{B}_k = [\boldsymbol{B}_k]_+^2$ (See Appendix B) and $[\boldsymbol{B}_k]_+$ is Hermitian, problem (41) is further equivalent to:

$$\begin{aligned}\frac{1}{2\lambda}\operatorname{tr}\left([\boldsymbol{B}_k]_+^2\right) - \frac{1}{4\lambda}\operatorname{tr}\left([\boldsymbol{B}_k]_+^2\right) &= \frac{1}{4\lambda}\operatorname{tr}\left([\boldsymbol{B}_k]_+^2\right)\\ &= \frac{1}{4\lambda}\left\|[\boldsymbol{B}_k]_+\right\|_F^2. \end{aligned} \quad (42)$$

Therefore, problem (34) is ultimately reformulated into the following form:

$$\inf_{\mu\geq 0}\left\{\mu\rho_2 + \frac{1}{4\lambda}\left\|[\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}]_+\right\|_F^2\right\} - \lambda\operatorname{tr}(\boldsymbol{R}_k^2), \quad (43)$$

substituting the results, problem (33) is ultimately equivalent to:

$$\inf_{\lambda\geq 0}\left\{\lambda\rho_1 + \frac{1}{K}\sum_{k=1}^{K}\left(\inf_{\mu\geq 0}\left\{\mu\rho_2 + \frac{1}{4\lambda}\left\|[\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}]_+\right\|_F^2\right\} - \lambda\operatorname{tr}(\boldsymbol{R}_k^2)\right)\right\}. \quad (44)$$

To improve computational efficiency, in the derivation, we aggregate the dual variables μ_k (originally independent for each sample $\boldsymbol{R}_k$) into a unified variable μ. The theoretical soundness of this approximation stems from the Aggregation Lemma in convex optimization theory. Specifically, when an optimization problem can be decomposed into a series of subproblems sharing an identical structure (i.e., the same constraint set $\mathcal{Z}_1$ and objective function form), the Lagrange multipliers $\{\mu_k\}$ associated with each sample in the dual problem exhibit aggregability. According to standard results in convex duality and constraint aggregation [22, Section 5.1.3], there exists a unified multiplier μ such that the objective value obtained from the aggregated problem forms a tight upper bound on the optimal value of the original problem. This implies that despite the introduced simplification, the solution we obtain retains a performance guarantee in the worst-case scenario, without underestimating the potential risk.

Therefore, based on this principle, when the constraints of the support set are taken into account, the objective function of the original optimization problem (7) is ultimately reformulated as follows:

$$\min_{\boldsymbol{w},\,\lambda\geq 0,\,\mu\geq 0}\left\{\lambda\rho_1 + \mu\rho_2 - \frac{\lambda}{K}\sum_{k=1}^{K}\operatorname{tr}(\boldsymbol{R}_k^2) + \frac{1}{4\lambda K}\sum_{k=1}^{K}\left\|\left[\,\boldsymbol{w}\boldsymbol{w}^H + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}\,\right]_+\right\|_F^2\right\}. \quad (45)$$

Physical Interpretation: The proposed optimization model (45) admits a clear physical interpretation. The regularization parameter λ adaptively balances the exploitation of the nominal sample information against the precaution for worst-case distributions. The dual variable μ acts as an energy threshold, intelligently filtering out extreme perturbations that are physically infeasible. The matrix positive-part operator $[\cdot]_+$ precisely identifies and retains only the interference subspace that poses a substantive threat to the beamformer's performance. Ultimately, by jointly optimizing the beamforming vector $\boldsymbol{w}$ and the dual variables λ and μ, the model achieves optimal immunity to the underlying statistical uncertainty with minimal performance degradation, all while strictly maintaining the desired array response constraint.

Thus, we have obtained the equivalent form (30) for the original optimization problem's objective function (7) when the support set constraint is ignored, as well as the equivalent form (45) that incorporates the support set constraint.

B. Dual Reformulation for the Constraints (Steering Vector)

In the following, we analyze the distributionally robust constraint on the steering vector in the original problem (7). This constraint ensures that the array response power remains no lower than a prescribed threshold even under the worst-case distributional mismatch of the steering vector. Analogously, by following the theoretical framework established previously for the covariance matrix uncertainty analysis and leveraging the strong duality of Wasserstein distributionally robust optimization (DRO), we derive the dual reformulation of the constraint in problem (7) as:

$$\sup_{\lambda\geq 0}\left(-\lambda\epsilon_a + \inf_{\boldsymbol{a}\in\mathcal{Z}_2}\left[|\boldsymbol{w}^H\boldsymbol{a}|^2 + \lambda\|\boldsymbol{a}-\boldsymbol{a}_0\|_2^2\right]\right) \geq \alpha. \quad (46)$$

where $\lambda \in \mathbb{R}$ and $\lambda \geq 0$ is the dual variable associated with the Wasserstein distance constraint. Analogously, we will discuss the two scenarios: ignoring the support set constraint and considering the support set constraint.

Case 1: Ignoring the Support Set (Unconstrained Approximation

When the support set $\mathcal{Z}_2$ is ignored, the inner minimization problem in problem (46) reduces to:

$$\inf_{\boldsymbol{a}\in\mathbb{C}^N}\left[|\boldsymbol{w}^H\boldsymbol{a}|^2 + \lambda\|\boldsymbol{a}-\boldsymbol{a}_0\|_2^2\right], \quad (47)$$

expanding the objective function of this problem yields:

$$\begin{aligned}&|\boldsymbol{w}^H\boldsymbol{a}|^2 + \lambda\|\boldsymbol{a}-\boldsymbol{a}_0\|_2^2\\ &= \boldsymbol{a}^H\boldsymbol{w}\boldsymbol{w}^H\boldsymbol{a} + \lambda\left(\|\boldsymbol{a}\|_2^2 - 2\Re\{\boldsymbol{a}^H\boldsymbol{a}_0\} + \|\boldsymbol{a}_0\|_2^2\right)\\ &= \boldsymbol{a}^H\left(\boldsymbol{w}\boldsymbol{w}^H + \lambda\boldsymbol{I}\right)\boldsymbol{a} - 2\lambda\Re\{\boldsymbol{a}^H\boldsymbol{a}_0\} + \lambda\|\boldsymbol{a}_0\|_2^2, \end{aligned} \quad (48)$$

since the quadratic form matrix $\boldsymbol{w}\boldsymbol{w}^H + \lambda\boldsymbol{I}$ is positive definite, the objective function is strictly convex with respect to $\boldsymbol{a}$, and the global minimum occurs at the point where the gradient vanishes. Thus, we set the Wirtinger gradient of the complex vector to zero:

$$\begin{aligned}\nabla_{\boldsymbol{a}}&\Big(\boldsymbol{a}^H\big(\boldsymbol{w}\boldsymbol{w}^H+\lambda\boldsymbol{I}\big)\boldsymbol{a}-2\lambda\Re\{\boldsymbol{a}^H\boldsymbol{a}_0\}+\lambda\|\boldsymbol{a}_0\|_2^2\Big)\\&=2\big(\boldsymbol{w}\boldsymbol{w}^H+\lambda\boldsymbol{I}\big)\boldsymbol{a}-2\lambda\boldsymbol{a}_0\\&=\boldsymbol{0},\end{aligned} \tag{49}$$

solving this yields the optimal solution:

$$\boldsymbol{a}^*=\lambda\big(\boldsymbol{w}\boldsymbol{w}^H+\lambda\boldsymbol{I}\big)^{-1}\boldsymbol{a}_0, \tag{50}$$

using the matrix inversion lemma, we expand $\big(\boldsymbol{w}\boldsymbol{w}^H+\lambda\boldsymbol{I}\big)^{-1}=\frac{1}{\lambda}\boldsymbol{I}-\frac{\boldsymbol{w}\boldsymbol{w}^H}{\lambda\left(\lambda+\|\boldsymbol{w}\|_2^2\right)}$. To simplify the closed-form solution, we note that the worst-case extremum of the array response $|\boldsymbol{w}^H\boldsymbol{a}|$ is achieved if and only if the Cauchy-Schwarz equality condition holds, i.e., the beamforming weight vector $\boldsymbol{w}$ is collinear with the nominal steering vector $\boldsymbol{a}_0$. Under this condition, we have $(\boldsymbol{w}\boldsymbol{w}^H)\boldsymbol{a}_0=\boldsymbol{w}(\boldsymbol{w}^H\boldsymbol{a}_0)=\|\boldsymbol{w}\|_2^2\boldsymbol{a}_0$. Substituting this into (50) yields the simplified result:

$$\boldsymbol{a}^*=\frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2}\,\boldsymbol{a}_0. \tag{51}$$

The optimal solution in (51) admits an intuitive physical interpretation: the robust steering vector $\boldsymbol{a}^*$ is a shrunken version of the nominal vector $\boldsymbol{a}_0$, scaled by the factor $\lambda/(\lambda+\|\boldsymbol{w}\|_2^2)\in(0,1]$. This scaling factor embodies an optimal balance between prior knowledge and learned robustness dictated by the distributionally robust framework.

Specifically, the regularization parameter $\lambda>0$, arising from the dual of the Wasserstein constraint, quantifies the "cost of uncertainty" or the perceived distributional mismatch. A larger λ (greater uncertainty) increases the scaling factor, pulling $\boldsymbol{a}^*$ closer to $\boldsymbol{a}_0$—the system becomes more conservative, relying more on prior information. Conversely, the beamformer's energy $\|\boldsymbol{w}\|_2^2$ reflects the design's aggressiveness. A more powerful beamformer (larger $\|\boldsymbol{w}\|_2^2$) further shrinks $\boldsymbol{a}^*$ away from $\boldsymbol{a}_0$, thereby reducing sensitivity to steering mismatches that could otherwise be amplified by high gain.

Therefore, the closed-form solution (51) reveals an intrinsic, data-driven trade-off: the framework automatically attenuates the steering vector in scenarios with high beamformer gain or low uncertainty tolerance, thereby guaranteeing robustness where it is most needed, while retaining the nominal direction when uncertainty is small or the beamformer is quiescent. This mechanism elucidates how distributional robustness translates into adaptive, physics-aware regularization.

Substituting the optimal solution (51) into (47), the inner minimization problem reduces to:

$$\frac{\lambda|\boldsymbol{w}^H\boldsymbol{a}_0|^2}{\lambda+\|\boldsymbol{w}\|_2^2}, \tag{52}$$

substituting this expression into the original dual problem (46) yields a single-variable optimization problem:

$$\sup_{\lambda\geq 0}\left(-\lambda\epsilon_a+\frac{\lambda\,|\boldsymbol{w}^H\boldsymbol{a}_0|^2}{\lambda+\|\boldsymbol{w}\|_2^2}\right), \tag{53}$$

since the objective function of this problem is strictly concave (its second derivative is negative definite), the global optimum occurs where its first derivative vanishes. Solving this yields the optimal dual variable:

$$\lambda^*=\|\boldsymbol{w}\|_2^2\cdot\left(\frac{|\boldsymbol{w}^H\boldsymbol{a}_0|}{\sqrt{\epsilon_a}\,\|\boldsymbol{w}\|_2}-1\right). \tag{54}$$

It is noted that a feasible solution exists only when the robust signal-to-interference-plus-noise ratio (RSNR) $\frac{|\boldsymbol{w}^H\boldsymbol{a}_0|}{\sqrt{\epsilon_a}\,\|\boldsymbol{w}\|_2}$ is greater than or equal to 1, i.e., $\lambda^*\geq 0$. This indicates that the nominal array response $|\boldsymbol{w}^H\boldsymbol{a}_0|$ must be at least sufficient to overcome the beamforming energy $\|\boldsymbol{w}\|_2$ amplified by the uncertainty radius $\sqrt{\epsilon_a}$. Only then does the system possess the capacity to afford the "uncertainty price" $\lambda^*>0$, thereby guaranteeing the worst-case performance through regularization. If $\mathrm{RSNR}<1$, it implies that either the uncertainty is too large or the nominal response is too weak. In this case, no regularization can satisfy the robustness requirement, rendering the original problem infeasible.

Substituting λ^* into problem (65), the worst-case minimum response power without support set constraints is finally obtained as:

$$\min_{G_2\in\mathcal{D}_2^w}\mathbb{E}\left[|\boldsymbol{w}^H\boldsymbol{a}|^2\right]=\big(|\boldsymbol{w}^H\boldsymbol{a}_0|-\sqrt{\epsilon_a}\,\|\boldsymbol{w}\|_2\big)^2, \tag{55}$$

combining the derived worst-case minimum response power with the original constraint that requires the array response power to be no lower than the threshold α in the worst-case scenario, the constraint in problem (7), when the support set constraint is ignored, is equivalent to the standard second-order cone (SOC) constraint :

$$|\boldsymbol{w}^H\boldsymbol{a}_0|\geq\sqrt{\alpha}+\sqrt{\epsilon_a}\,\|\boldsymbol{w}\|_2. \tag{56}$$

Case 2: Considering the Support Set (Constrained Optimization)

Next, we derive the transformed form of the original dual problem (46) when the support set is taken into account. Consider the support set defined as $\mathcal{Z}_2=\{\boldsymbol{a}\in\mathbb{C}^N\mid(1-\Delta)N\leq\|\boldsymbol{a}\|_2^2\leq(1+\Delta)N\}$, where $0<\Delta<1$ is the relative fluctuation coefficient of the norm, determined by the amplitude-phase error level of the practical array system(typically $0.05\leq\Delta\leq 0.3$ in engineering applications).

First, based on the unconstrained optimal solution $\boldsymbol{a}^*$ in (51), it is evident that for any $\lambda\geq 0$, we have:

$$\|\boldsymbol{a}^*\|_2^2=\left(\frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2}\right)^2\|\boldsymbol{a}_0\|_2^2=\left(\frac{\lambda}{\lambda+\|\boldsymbol{w}\|^2}\right)^2N<N, \tag{57}$$

therefore, for the unconstrained optimal solution $\boldsymbol{a}^*$, the upper-bound constraint in the support set constraint naturally holds ($\|\boldsymbol{a}^*\|_2^2<N<(1+\Delta)N$). Consequently, the constrained optimal solution always lies strictly within the interior of the upper bound. Thus, the upper-bound constraint is never active at the optimal solution.We only need to consider the relationship between the unconstrained optimal solution $\boldsymbol{a}^*$ and the lower bound of the support set constraint.

When $\|\boldsymbol{a}^*\|_2^2<(1-\Delta)N$, the global optimum of the strictly concave function (48) cannot be attained in the interior of the support set $\mathcal{Z}_2$; instead, it must lie on the boundary

$\|\boldsymbol{a}\|_2^2 = (1-\Delta)N$, meaning the lower-bound constraint becomes active. In this case, the inner minimization problem of (46) becomes a convex optimization problem with an equality constraint:

$$\inf_{\|\boldsymbol{a}\|_2^2=(1-\Delta)N} \left[|\boldsymbol{w}^H\boldsymbol{a}|^2 + \lambda\|\boldsymbol{a}-\boldsymbol{a}_0\|_2^2\right], \tag{58}$$

we introduce a non-negative dual variable $\mu \geq 0$ to construct the Lagrangian function corresponding to the lower-bound constraint:

$$\begin{aligned}\mathcal{L}(\boldsymbol{a},\mu) = \boldsymbol{a}^H\big(\boldsymbol{w}\boldsymbol{w}^H + \lambda\mathbf{I}\big)\boldsymbol{a} - 2\lambda\Re\{\boldsymbol{a}^H\boldsymbol{a}_0\} \\ + \lambda\|\boldsymbol{a}_0\|_2^2 - \mu\Big(\|\boldsymbol{a}\|_2^2 - (1-\Delta)N\Big),\end{aligned} \tag{59}$$

according to the KKT conditions of convex optimization, the optimal solution $\boldsymbol{a}_c^*$ ($\boldsymbol{a}_c^*$ denote the optimal solution when the support set lower-bound constraint is active, i.e., when $\|\boldsymbol{a}\|_2^2 = (1-\Delta)N$, to distinguish it from the $\boldsymbol{a}^*$ in (51) which was obtained by ignoring the support set constraint) must satisfy that its gradient vanishes:

$$\nabla_{\boldsymbol{a}}\mathcal{L}(\boldsymbol{a}_c^*,\mu) = 2\Big(\boldsymbol{w}\boldsymbol{w}^H + (\lambda-\mu)\boldsymbol{I}\Big)\boldsymbol{a}_c^* - 2\lambda\boldsymbol{a}_0 = \mathbf{0}, \tag{60}$$

similar to the treatment for the optimal solution without support set constraints, simplification yields:

$$\boldsymbol{a}_c^* = \frac{\lambda}{\lambda + \|\boldsymbol{w}\|_2^2 - \mu}\,\boldsymbol{a}_0, \tag{61}$$

this optimal solution remains a scalar scaling of the nominal steering vector $\boldsymbol{a}_0$, which perfectly matches the physical scenario of array amplitude-phase errors. The dual μ variable at this point satisfies the non-negativity requirement of the KKT conditions (See Appendix C).

Substituting this optimal solution into the original problem (46), the inner minimization function simplifies to:

$$\begin{aligned}\left|\boldsymbol{w}^H\boldsymbol{a}_c^*\right|^2 + \lambda\left\|\boldsymbol{a}_c^* - \boldsymbol{a}_0\right\|_2^2 = \left|\boldsymbol{w}^H \cdot \frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2-\mu}\boldsymbol{a}_0\right|^2 \\ + \lambda\left\|\frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2-\mu}\boldsymbol{a}_0 - \boldsymbol{a}_0\right\|_2^2.\end{aligned} \tag{62}$$

Combining the boundary constraint $\|\boldsymbol{a}^*\|_2^2 = (1-\Delta)N$ with (61) yields:

$$\frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2-\mu} = \sqrt{1-\Delta}, \tag{63}$$

substituting this expression into (62), equation (62) reduces to:

$$\big(1-\Delta\big)\left|\boldsymbol{w}^H\boldsymbol{a}_0\right|^2 + \lambda C, \tag{64}$$

where $C = N\big(2-2\sqrt{1-\Delta}-\Delta\big)$, which is a constant depending solely on N and Δ. Substituting this expression into the original problem (46) yields a single-variable optimization problem:

$$\sup_{\lambda\geq 0}\Big((1-\Delta)\left|\boldsymbol{w}^H\boldsymbol{a}_0\right|^2 + \lambda\big(C-\epsilon_a\big)\Big), \tag{65}$$

this is a linear function in λ, and the value of its global supremum is determined by two cases:

1. When $C-\epsilon_a > 0$, $g(\lambda)$ increases monotonically with λ, and its supremum is $+\infty$, which has no physical meaning.

2. When $C-\epsilon_a \leq 0$, $g(\lambda)$ decreases monotonically with λ, and the global supremum is attained at $\lambda^* = 0$, corresponding to the worst-case scenario. In this case, since the optimal dual variable is $\lambda^* = 0$, substituting it back yields the global supremum of the dual problem as $(1-\Delta)\left|\boldsymbol{w}^H\boldsymbol{a}_0\right|^2$. Combining this with the constraint in the original problem (7), the supremum of the dual problem must satisfy $(1-\Delta)\left|\boldsymbol{w}^H\boldsymbol{a}_0\right|^2 \geq \alpha$. Therefore, the final equivalent convex constraint under the support set constraint is obtained as:

$$|\boldsymbol{w}^H\boldsymbol{a}_0| \geq \frac{\sqrt{\alpha}}{\sqrt{1-\Delta}}. \tag{66}$$

Activation Criterion for the Support Set Constraint: Next, we derive the quantitative relationship between the activation condition of the lower-bound constraint and the error parameters. The necessary condition for the lower-bound constraint to be active is that the squared norm of the unconstrained optimal solution $\boldsymbol{a}^*$ is less than the lower bound of the support set, i.e., $\|\boldsymbol{a}^*\|_2^2 < (1-\Delta)N$. Substituting the unconstrained optimal solution $\boldsymbol{a}^*$ from (51) yields:

$$\left(\frac{\lambda}{\lambda+\|\boldsymbol{w}\|_2^2}\right)^2 N < (1-\Delta)N, \tag{67}$$

after simplification, we obtain:

$$\lambda < \frac{\sqrt{1-\Delta}}{1-\sqrt{1-\Delta}}\|\boldsymbol{w}\|_2^2, \tag{68}$$

in this inequality, λ is the dual variable in the original dual problem. Under the worst-case analysis, this dual variable takes its optimal value λ^*, which attains the supremum of the original dual problem. Substituting the optimal dual variable λ^* from equation (54) into the inequality and simplifying yields:

$$\frac{|\boldsymbol{w}^H\boldsymbol{a}_0|}{\sqrt{\epsilon_a}\|\boldsymbol{w}\|_2} < 1 + \frac{\sqrt{1-\Delta}}{1-\sqrt{1-\Delta}}, \tag{69}$$

it can be seen that the necessary condition for the lower-bound constraint to be active is that the RSNR is less than a threshold that depends solely on the array amplitude-phase error, i.e., $1+\frac{\sqrt{1-\Delta}}{1-\sqrt{1-\Delta}}$.

On the other hand, for the lower-bound constraint to be active, it must also satisfy the condition that the slope of the linear function in λin problem (65) is non-positive, i.e., $C-\epsilon_a \leq 0$. Consequently, we obtain:

$$N\Big(2-2\sqrt{1-\Delta}-\Delta\Big) \leq \epsilon_a, \tag{70}$$

in practice, however, this inequality may not hold.

Based on the derivation above, the activation of the lower-bound constraint of the support set does not always occur. It is jointly determined by two computable criteria: equation (69) measures the relative robustness of the nominal system performance against uncertainty, while equation (70) compares the physical adjustment cost with the statistical uncertainty cost. The optimal solution will lie on the boundary $\|\boldsymbol{a}\|_2^2 = (1-\Delta)N$ if and only if both conditions are satisfied simultaneously. In this case, the constraint (66) after the activation of the lower bound should be adopted to reduce the conservativeness inherent in the unconstrained (or support-set-ignoring) case and improve performance. This criterion provides crucial guidance for system design: by estimating or measuring the error parameters (ϵ_a, Δ) and calculating the RSNR, one can pre-determine whether the lower-bound constraint of the support set needs to be considered, thereby selecting the appropriate final constraint.

C. Unified Distributionally Robust Optimization Problem

SOLVING VIA SDP-SOCP RELAXATION

Owing to the presence of the outer-product matrix $\boldsymbol{w}\boldsymbol{w}^H$ and the positive-part operator $[\cdot]_+$, problem (45) is non-convex in its entirety. We therefore begin with a variable lifting step, defining $\boldsymbol{W} = \boldsymbol{w}\boldsymbol{w}^H$. By first ignoring the rank-one constraint $\mathrm{rank}(\boldsymbol{W}) = 1$, we obtain the relaxed variable constraint $\boldsymbol{W} \succeq 0$. To handle the positive-part norm term, we introduce the following theorem:

Lemma 1: For any Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, the squared Frobenius norm of its positive semidefinite projection admits the following equivalent variational form:

$$\left\|[\boldsymbol{A}]_+\right\|_F^2 = \min_{\boldsymbol{Y}\succeq \boldsymbol{0},\, \boldsymbol{Y}\succeq \boldsymbol{A}} \|\boldsymbol{Y}\|_F^2. \tag{71}$$

Proof 1: This variational characterization of the squared Frobenius norm of the positive semidefinite projection follows directly from the general theory of projection onto convex cones. A complete exposition can be found in [22, Section 8.1.1] for the Euclidean projection onto the semidefinite cone, and in [24, Section 2.6] for a broader discussion of proximal operators and cone projections.

According to this lemma, by introducing auxiliary variables $\boldsymbol{Y}_k \succeq 0$, the positive-part norm term in problem (45) is equivalent to (as shown earlier, we also define $\boldsymbol{B}_k = \boldsymbol{W} + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}$):

$$\|[\boldsymbol{B}_k]_+\|_F^2 = \min_{\boldsymbol{Y}_k\succeq 0,\, \boldsymbol{Y}_k\succeq \boldsymbol{B}_k} \|\boldsymbol{Y}_k\|_F^2\,. \tag{72}$$

Therefore, this term in problem (45) can be precisely replaced with the auxiliary variable $\boldsymbol{Y}_k$ in its minimization form. Since the term $\frac{1}{\lambda}\|\boldsymbol{Y}_k\|_F^2$ is non-convex with respect to the optimization variables λ and $\boldsymbol{Y}_k$, to convert it into a convex form, we introduce an auxiliary variable $s_k \geq 0$ and add the following constraint:

$$\frac{1}{\lambda}\|\boldsymbol{Y}_k\|_F^2 \leq s_k. \tag{73}$$

Moreover, since this constraint is equivalent to $\|\boldsymbol{Y}_k\|_F^2 \leq \lambda s_k$, which is a convex rotated second-order cone (RSOC) constraint, (73) can be equivalently expressed as a second-order cone (SOC) constraint:

$$\left\|\begin{bmatrix} 2\,\mathrm{vec}_{\mathbb{R}}(\boldsymbol{Y}_k) \\ \lambda - s_k \end{bmatrix}\right\|_2 \leq \lambda + s_k. \tag{74}$$

Here, $\mathrm{vec}_{\mathbb{R}}(\cdot) : \mathbb{C}^{m\times n} \to \mathbb{R}^{2mn}$ denotes the real vectorization operator for a complex matrix. It is defined by first decomposing the matrix Z into its real part $\Re(\boldsymbol{Z})$ andmaginary parts $\Im(\boldsymbol{Z})$, and then concatenating their vectorized forms into a single real-valued column vector, i.e., $\mathrm{vec}_{\mathbb{R}}(\boldsymbol{Z}) = [\mathrm{vec}(\Re(\boldsymbol{Z}))^\top, \mathrm{vec}(\Im(\boldsymbol{Z}))^\top]^\top$.

Thus, with the rank-one constraint relaxed, problem (45) has been converted into a convex form, yielding the final, solvable semidefinite relaxation (SDR) problem:

$$\begin{aligned}
\min_{\substack{\boldsymbol{W},\lambda,\mu,\\ \{\boldsymbol{Y}_k\},\{s_k\}}} \quad & \lambda\rho_1 + \mu\rho_2 - \frac{\lambda}{K}\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{R}_k^2\right) + \frac{1}{4K}\sum_{k=1}^{K} s_k \\
\text{s.t.} \quad & \boldsymbol{W} \succeq 0, \\
& \lambda \geq 0, \quad \mu \geq 0, \\
& \boldsymbol{Y}_k \succeq 0, \quad k = 1,\ldots,K, \\
& \boldsymbol{Y}_k \succeq \boldsymbol{W} + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}, \quad k = 1,\ldots,K, \\
& \left\|\begin{bmatrix} 2\,\mathrm{vec}_{\mathbb{R}}(\boldsymbol{Y}_k) \\ \lambda - s_k \end{bmatrix}\right\|_2 \leq \lambda + s_k, \quad s_k \geq 0, \quad k = 1,\ldots,K,
\end{aligned} \tag{75}$$

this problem constitutes a convex mixed semidefinite-second-order cone programming (SDP-SOCP) problem. It can therefore be solved efficiently using standard convex optimization solvers, such as Mosek or CVX.

RANK-ONE RECOVERY VIA DIFFERENCE OF CONVEX FUNCTIONS ALGORITHM (DCA)

Since problem (75) ignores the rank-one constraint, we next proceed to obtain a rank-one solution to the problem by introducing a penalty function and solving iteratively. We first consider the penalty function $P(\boldsymbol{W}) = \mathrm{tr}(\boldsymbol{W}) - \lambda_{\max}(\boldsymbol{W})$. This penalty is a standard difference of convex (DC) function, which can be decomposed as $P(W) = p_1(W) - p_2(W)$, where $p_1(W) = \mathrm{tr}(W)$ is convex and $p_2(W) = \lambda_{\max}(W)$ is also convex on the positive semidefinite cone. Since $-\lambda_{\max}(\boldsymbol{W})$ is non-convex, the penalty function $P(\boldsymbol{W})$ is non-convex overall. This non-convexity can lead to local optima traps or convergence instability in the optimization process, making it difficult to guarantee the global optimality of the rank-one solution.Therefore, by replacing the penalty function via a Rayleigh quotient-driven convex upper bound, we reconstruct the penalty term, ensuring the upper-bound property of the objective function while preserving optimality guarantees. This convex surrogate construction follows the core step of DCA: subgradient linearization of the convex term in the DC decomposition.

Specifically, we replace $P(\boldsymbol{W})$ with a conservative convex upper bound $G^{(t)}(\boldsymbol{W})$, whose form and theoretical justification are as follows:

$$G^{(t)}(\boldsymbol{W}) = \mathrm{tr}(\boldsymbol{W}) - \left(\boldsymbol{u}_1^{(t)}\right)^H \boldsymbol{W}\,\boldsymbol{u}_1^{(t)}, \tag{76}$$

where the superscript (t) denotes the t-th iteration, and $\boldsymbol{u}_1^{(t)}$ is the estimated principal unit eigenvector computed from the solution of the previous iteration, $\boldsymbol{W}^{(t-1)}$. For this penalty function, the conservativeness is strictly guaranteed by the upper-bound property of eigenvalues in the Rayleigh quotient: for any unit vector $\boldsymbol{u}_1^{(t)}$, we have $\boldsymbol{u}_1^{(t)H}\boldsymbol{W}\boldsymbol{u}_1^{(t)} \leq \lambda_{\max}(\boldsymbol{W})$. Rearranging the terms yields $G^{(t)}(\boldsymbol{W}) \geq P(\boldsymbol{W})$. Moreover, $G^{(t)}(\boldsymbol{W})$ is a convex function because it consists of a linear term and a trace term. Therefore, since $G^{(t)}(\boldsymbol{W})$ is an upper bound of $P(\boldsymbol{W})$, the penalty function (76) essentially minimizes the sum of the non-principal eigenvalues, $\sum_{i=2}^{N}\lambda_i$.

When this sum approaches zero, the matrix naturally collapses to a rank-one matrix, but also, due to its convexity, guarantees convergence to a stationary point of the penalized problem. This provides both theoretical and algorithmic support for the stable realization of rank-one recovery. The convergence property is rigorously guaranteed by classical DC programming theory [25].

Therefore, via the DCA convex surrogate (76), we can iteratively obtain a rank-one solution to problem (75). In the t-th DCA iteration of this procedure, we solve the following problem by introducing a penalty term (which is adaptively updated based on the solution from the previous step to enforce the rank-one constraint):

$$\begin{aligned}
\min_{\substack{\boldsymbol{W},\lambda,\mu,\\ \{\boldsymbol{Y}_k\},\{s_k\}}} \quad & \lambda\rho_1 + \mu\rho_2 - \frac{\lambda}{K}\sum_{k=1}^{K}\operatorname{tr}\left(\boldsymbol{R}_k^2\right) + \frac{1}{4K}\sum_{k=1}^{K} s_k \\
& + \xi\left(\operatorname{tr}(\boldsymbol{W}) - \left(\boldsymbol{u}_1^{(t)}\right)^H \boldsymbol{W}\boldsymbol{u}_1^{(t)}\right) \\
\text{s.t.} \quad & \boldsymbol{W} \succeq 0, \\
& \lambda \geq 0,\ \mu \geq 0, \\
& \boldsymbol{Y}_k \succeq 0, \quad k = 1,\ldots,K, \\
& \boldsymbol{Y}_k \succeq \boldsymbol{W} + 2\lambda\boldsymbol{R}_k - \mu\boldsymbol{I}, \quad k = 1,\ldots,K, \\
& \left\| \begin{bmatrix} 2\operatorname{vec}_{\mathbb{R}}(\boldsymbol{Y}_k) \\ \lambda - s_k \end{bmatrix} \right\|_2 \leq \lambda + s_k,\ s_k \geq 0,\ k = 1,\ldots,K.
\end{aligned} \tag{77}$$

Here, ξ denotes a predefined penalty factor, and a sufficiently large value should be chosen to enforce the rank-one constraint. The high-rank solution $\boldsymbol{W}^*$ obtained from problem (75) serves as the initial point $\boldsymbol{W}^{(0)}$ for the iterative procedure in problem (77).

IV. Data-Driven Parameter Calibration via Statistical Robustness Guarantees

In the previous chapter, we transformed the distributionally robust beamforming problem into a tractable convex optimization form. Its performance and robustness are ultimately governed by two key sets of parameters: the Wasserstein radii (ρ_1 and ϵ_a, quantifying the distributional uncertainty) and the support set parameters (Δ, ρ_2, quantifying the physical amplitude-phase errors and ensuring physical realizability). However, a fundamental question remains unresolved: how should these parameters be selected in engineering practice? Traditional methods, such as diagonal loading, rely on empirical tuning, lack a theoretical foundation, and fail to provide a provably optimal trade-off between performance and robustness.

If the parameters are chosen too conservatively (with excessively large values), it will lead to unnecessary performance degradation of the beamformer; if they are too optimistic (with excessively small values), robustness in the worst-case scenario cannot be guaranteed. To address this challenge, this chapter, based on the finite-sample theory of distributionally robust optimization established by Blanchet et al. [17], proposes a data-driven parameter calibration framework with rigorous statistical guarantees. The core idea of this framework is to leverage available sample data (e.g., training snapshots) to guarantee, with high probability, that the true distributions of the interference-plus-noise covariance matrix and the steering vector lie within the Wasserstein uncertainty sets we construct. Thereby, the parameter selection problem is transformed into an optimization problem that meets statistical confidence requirements.

Specifically, this chapter first elaborates, based on the chi-square fitting criterion, the calibration procedure for the Wasserstein radius ρ_1 and the support-set parameter ρ_2, ensuring that they cover the true distributions with a preset confidence level. Next, it presents calibration methods for the Wasserstein radius ϵ_a under different scenarios. Finally, through the design in this chapter, the key parameters are completely freed from empirical tuning, realizing a paradigm shift from "manual parameter setting" to "data-driven parameter determination," which lays a solid foundation for the reliable application of the proposed algorithm in practical systems.

We first employ a three-step quantile estimation procedure to calibrate the parameter ρ_1. This entire process is performed without requiring any prior assumptions about the underlying distribution of the covariance matrices:

1.**Data Preprocessing via Sliding Window and Bootstrap:** To generate the set of sample covariance matrices $\{\widehat{\boldsymbol{R}}^{(1)},\ldots,\widehat{\boldsymbol{R}}^{(B)}\}$ required for quantile estimation, we employ a two-stage processing scheme. This approach efficiently utilizes the limited temporal data while accounting for their inherent statistical properties.

a) Stage 1: Sliding Window Processing for Correlated Samples: For an array with N elements, suppose a total of T temporal snapshots $\boldsymbol{x}_t \in \mathbb{C}^N$ $(t = 1,\ldots,T)$ are available. First, a sliding window of length L is employed to partition the data into K overlapping blocks, where the window length satisfies $L \geq 2N$ (to ensure the empirical covariance matrix is well-conditioned and to avoid ill-conditioning issues). Starting from the first snapshot at $t = 1$, the k-th window ($k = 1,\ldots,K$, with the total number of windows $K = T - L + 1$) covers snapshots k through $k + L - 1$. The corresponding sample covariance matrix is computed as:

$$\tilde{\boldsymbol{R}}_k = \frac{1}{L}\sum_{t=k}^{k+L-1} \boldsymbol{x}_t\boldsymbol{x}_t^H, \tag{78}$$

this yields the initial set of covariance matrices $\{\widetilde{\boldsymbol{R}}_1,\ldots,\widetilde{\boldsymbol{R}}_K\}$. This collection of matrices maintains spatial validity but exhibits temporal correlation, thereby maximizing the utilization value of the limited observational data.

b) Stage 2: Bootstrap Resampling for Generating I.I.D. Pseudo-Observations: Quantile estimation requires input samples to satisfy the independent and identically distributed (i.i.d.) condition. Therefore, for the initial sample set described above, which exhibits correlation, we employ a nonparametric bootstrap resampling algorithm to generate i.i.d. pseudo-observations. Set the total number of bootstrap trials to B (e.g., $B = 2000$). The procedure for a single trial is as follows:

1) Resampling: From the initial set $\{\widetilde{\boldsymbol{R}}_1,\ldots,\widetilde{\boldsymbol{R}}_K\}$, draw K samples uniformly with replacement to form the b-th bootstrap sample set $\{\widetilde{\boldsymbol{R}}_1^{(b)},\ldots,\widetilde{\boldsymbol{R}}_K^{(b)}\}$.

2) Aggregation: Compute the mean of this bootstrap sample set to obtain the b-th pseudo-observation matrix:

$$\widehat{\boldsymbol{R}}^{(b)} = \frac{1}{K}\sum_{i=1}^{K} \widetilde{\boldsymbol{R}}_i^{(b)}. \tag{79}$$

The resulting sample set $\{\widehat{\boldsymbol{R}}^{(1)}, \ldots, \widehat{\boldsymbol{R}}^{(B)}\}$ contains B independent and identically distributed (i.i.d.) samples. The bootstrap mean operation provides a smoothing and noise-reduction effect, while the resampling mechanism effectively mitigates the temporal correlation introduced by the sliding window. This process constructs a robust empirical distribution for the subsequent quantile estimation.

2. **Distance Computation:** Calculate the empirical squared Wasserstein-2 distance for each sample $\widehat{\boldsymbol{R}}^{(b)}$ relative to the reference center $\boldsymbol{S}_0$. For covariance matrices, the squared Frobenius norm of their difference is employed as an efficient approximation of the Wasserstein-2 distance. This metric offers both low computational complexity and statistical consistency:

$$d^{(b)} = \left\|\boldsymbol{R}^{(b)} - \boldsymbol{S}_0\right\|_F^2 \quad i = 1, \ldots, B. \tag{80}$$

3. **Radius Determination:** Given a target confidence level $1-\alpha$ (e.g., $\alpha = 0.05$ corresponds to a 95% confidence level), set the Wasserstein radius ρ_1 to the upper α-quantile of the empirical distance set $\{d^{(1)}, \ldots, d^{(B)}\}$:

$$\rho_1(\alpha) = \text{Quantile}_{1-\alpha}(\{d_1, \ldots, d_B\}). \tag{81}$$

This value represents the minimum radius that satisfies the constraint, guaranteeing that at least a $1-\alpha$ proportion of the observed samples lie within the Wasserstein uncertainty ball $\mathcal{D}_1^w(\rho_1(\alpha))$.

Similarly, we can leverage the bootstrap sample set $\{\widehat{\boldsymbol{R}}^{(1)}, \ldots, \widehat{\boldsymbol{R}}^{(B)}\}$ to determine the parameter ρ_2, such that the trace of the true covariance matrix $\text{tr}(\boldsymbol{R})$ is guaranteed, with high probability, not to exceed this value. Specifically, for each sample, compute its trace $t_b = \text{tr}(\boldsymbol{R}_b)$, for $b = 1, \ldots, B$. Then, choose a high confidence level $1-\alpha$, and set ρ_2 to the $(1-\alpha)$ empirical quantile of the sample trace set $\{t_b\}$:

$$\rho_1(\alpha) = \text{Quantile}_{1-\alpha}(\{t_1, \ldots, t_B\}). \tag{82}$$

This ensures that at least $100(1-\alpha)\%$ of the training samples have a power level below ρ_2. Within the distributionally robust framework, this is equivalent to assuming that the support of the true distribution G_1 is contained within the set defined by the trace constraint ρ_2, and this assumption holds with an empirical probability of $1-\alpha$.

Next, we derive the determination of the parameter ϵ_a. If the error distribution of the steering vector (e.g., from DOA estimation error, array calibration error) is explicitly known, a similar three-step procedure can be employed to determine a high-confidence value for ϵ_a. However, in practical scenarios, the error distribution of the steering vector is often difficult to obtain. In cases where an absolute bound can be determined from prior knowledge (e.g., sensor calibration accuracy, the maximum allowable error of the angle estimation module), we propose a conservative estimation method based on the worst-case bound to provide a rigorous robustness guarantee.

Specifically, assume the absolute upper bound for the Direction-of-Arrival (DOA) estimation error is known to be $\Theta_{\max}$, and the magnitude upper bound for the diagonal perturbation matrix $\boldsymbol{\Gamma}$ (satisfying $\boldsymbol{\Gamma}^H\boldsymbol{\Gamma} \preceq \boldsymbol{I}$) representing array amplitude-phase errors is $\Delta_{\max}$. Under these premises, the parameter ϵ_a corresponding to the worst-case deviation of the steering vector $\boldsymbol{a}$ from its nominal value $\boldsymbol{a}_0$ (corresponding to the nominal DOA θ_0) can be obtained by solving the following optimization problem:

$$\epsilon_a^{\max} = \max_{|\delta\theta|\le\Theta_{\max},\ \|\boldsymbol{\Gamma}\|\le\Delta_{\max}} \left\|\left(\boldsymbol{I}+\Delta_{\max}\boldsymbol{\Gamma}\right)\cdot\boldsymbol{a}\left(\theta_0+\delta\theta\right)-\boldsymbol{a}_0\right\|_2^2, \tag{83}$$

we set the Wasserstein radius to this maximum value, i.e., $\epsilon_a = \epsilon_a^{\max}$. This method ensures that, within the given physical error bounds, the true steering vector falls into the constructed Wasserstein-ball uncertainty set $\mathcal{D}_2^w$ with 100% probability. Its advantage lies in requiring no statistical assumptions about the error distribution and providing the strongest robustness guarantee. However, since it targets an extreme and low-probability joint error scenario, the resulting ϵ_a is typically large, which may cause the beamforming design to become overly conservative, thereby sacrificing the system's average performance to some extent.

V. Simulation Results

For numerical simulations, we consider a uniform linear array (ULA) of $N = 10$ omnidirectional sensors with half-wavelength spacing. The additive noise is modeled as a spatially and temporally white complex Gaussian process with zero mean and unit variance (0 dB). Two uncorrelated interfering sources are present, each with an interference-to-noise ratio (INR) of 30 dB, arriving from directions $-20°$ and $30°$. The true direction of arrival (DOA) of the desired signal is $\theta_{\text{true}} = 3°$, and the desired signal is always present in the training data. Unless otherwise stated, the number of training snapshots is $T = 100$, and all performance metrics are averaged over 200 independent Monte Carlo trials. The specific steering vector mismatch scenarios and the compared algorithms are detailed in the following subsections. The optimization problems are solved using the CVX toolbox with the default SDPT3 solver.

For the proposed Wasserstein-2 DRO framework, the parameters ρ_1 and ρ_2 for the covariance ambiguity set are calibrated via the bootstrap-based quantile estimation procedure described in Section IV, using a sliding window of length $L = 3N = 30$, $B = 2000$ bootstrap trials, and a 95% confidence level ($\alpha = 0.05$). The steering vector uncertainty radius is set to $\epsilon_a = 3.0$, corresponding to the squared ℓ_2-norm of the mismatch vector for a nominal $2°$ error. For the DCA procedure in the support-constrained variants, the penalty factor is scaled as $\xi = 100/\beta$ with $\beta = \text{tr}(\boldsymbol{S}_0)/N$ for numerical stability, and iterations stop when the relative eigenvalue gap falls below 10^{-4} or after 5 iterations. The steering vector support parameter is $\Delta = 0.1$ for the double-support variant.

For performance comparison, the proposed methods (the support-free DRO, the covariance-support DRO, and, where applicable, the double-support DRO) are evaluated alongside several representative robust adaptive beamforming algorithms. The standard Capon beamformer (SCB) serves as a baseline without any robustness mechanism [5], while diagonal loading (DL) with a fixed loading level of 15 dB represents a widely used engineering heuristic [7]. The robust Capon beamformer (RCB) of Li, Stoica, and Wang [11] is implemented with its spherical uncertainty set radius set to $\epsilon = 3.0$, matching the nominal 2° mismatch. The distributionally robust MVDR beamformer (DR-MVB) of B. Li, Rong, Sun, and Teo [14] is configured as in the original work, with the error variance $\sigma^2 = 0.1$ and the probability threshold $p = 0.9$. The moment-based DRO beamformer (M-DRO) of Irani et al. [15] adopts the parameter settings specified therein, namely $\rho_1 = 0.001\|\mathbf{S}_0\|_F$, $\rho_2 = 1.1\,\mathrm{tr}(\mathbf{S}_0)$, $\gamma_1 = 0.01\|\mathbf{a}_0\|$, $\gamma_2 = 0.1$, $\Delta = 0.1$, and the penalty coefficient $\alpha = 10^3$. For this method, the angular sector $[0^\circ, 6^\circ]$ centered at the true DOA (3°) is used for moment estimation, which provides the beamformer with accurate prior information regarding the signal direction.

Fig. 1 illustrates the output SINR versus input SNR for all considered methods under a 2° angle mismatch. The standard Capon beamformer (SCB) [5] suffers severe signal self-nulling at high SNR. Diagonal loading (DL, 15 dB) [7] provides limited robustness, with performance declining beyond 10 dB as its fixed loading becomes inadequate. The robust Capon beamformer (RCB) [11] performs well since its uncertainty radius ($\epsilon = 3.0$) perfectly matches the mismatch. The distributionally robust MVDR beamformer (DR-MVB) of B. Li et al. [14] offers limited gain at high SNR as it ignores covariance uncertainty. Crucially, even when configured with the ideal angular sector (centered at the true DOA), the moment-based DRO beamformer (M-DRO) of Irani et al. [15] exhibits significant performance degradation, demonstrating the inherent conservatism of its ambiguity sets. In the low SNR regime, the proposed methods perform comparably to, and in some cases slightly below, the competing algorithms, as the data-driven calibration may introduce a marginally higher regularization level in noise-dominated scenarios. However, as SNR increases, the advantages of the data-driven, Wasserstein-2 DRO framework become decisive. Both proposed variants outperform all benchmarks. The covariance-support DRO consistently surpasses the support-free version, achieving an output SINR exceeding 40 dB at 50 dB input SNR—a new state-of-the-art for non-reconstruction-based robust beamforming. This result demonstrates that by integrating physical support set constraints and adaptive parameter calibration, the proposed framework can effectively alleviate the excessive conservatism of moment-based DRO algorithms, thereby fully verifying the efficacy of the presented method.

Fig. 2 presents the corresponding results under a 4° angle mismatch. The standard Capon beamformer [5] collapses even earlier as the larger mismatch exacerbates signal self-nulling. Diagonal loading (15 dB) [7] similarly degrades beyond 10 dB and cannot recover. The robust Capon beamformer [11], whose fixed uncertainty radius ($\epsilon = 3.0$) was optimal for the 2° case, now suffers severe performance loss because the actual mismatch significantly exceeds the assumed uncertainty set, highlighting the sensitivity of fixed-prior methods to parameter

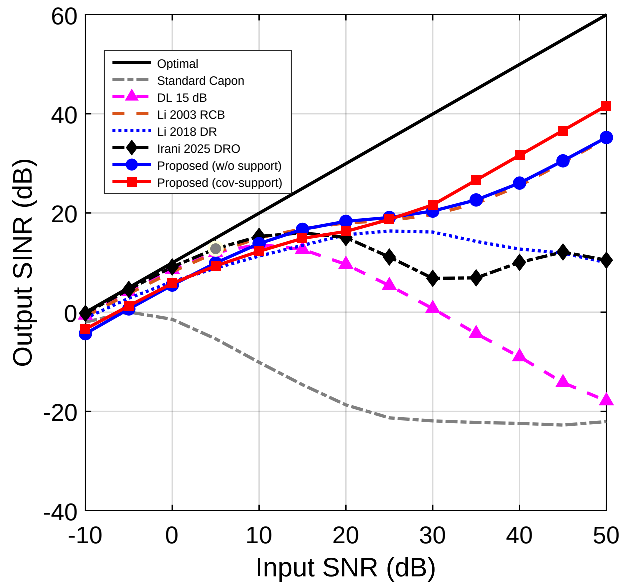

Fig. 1. Performance comparison under 2° angle mismatch.

mismatch. The distributionally robust MVDR beamformer [14] provides limited improvement as it only addresses the steering vector uncertainty. The moment-based DRO beamformer [15] exhibits substantial degradation, further confirming the inherent conservatism of its ambiguity sets. In contrast, both proposed variants maintain robust performance—without any parameter adjustment—as the data-driven calibration automatically adapts the regularization to the increased mismatch. The covariance-support DRO again outperforms the support-free version, and at 50 dB SNR, its output SINR remains approximately 35 dB, far exceeding all competing methods. This sustained advantage confirms that the physically motivated trace support constraint, by capping the worst-case INC matrix power, effectively prevents the excessive diagonal loading that would otherwise arise from the enlarged data-driven radius under large mismatches, thereby preserving the beamformer's interference suppression capability at high SNR. These results highlight the critical merit of the proposed data-driven framework: while fixed-prior and moment-based approaches fail to generalize beyond their design conditions, the Wasserstein-2 strategy with bootstrap-based calibration preserves near-optimal performance across different mismatch levels without any manual re-tuning.

Fig. 3 illustrates the output SINR versus angle mismatch for all considered methods at a fixed input SNR of 20 dB. The angle mismatch is swept from 1° to 6°. The standard Capon beamformer (SCB) [5] is highly sensitive to steering vector errors and its performance collapses immediately even at a 1° mismatch, demonstrating the well-known signal self-nulling phenomenon. Diagonal loading (15 dB) [7] provides a notable improvement in robustness over SCB; however, its performance degrades steadily as the angle mismatch grows, revealing that a fixed loading factor is ultimately insufficient to compensate for increasingly severe steering vector errors.

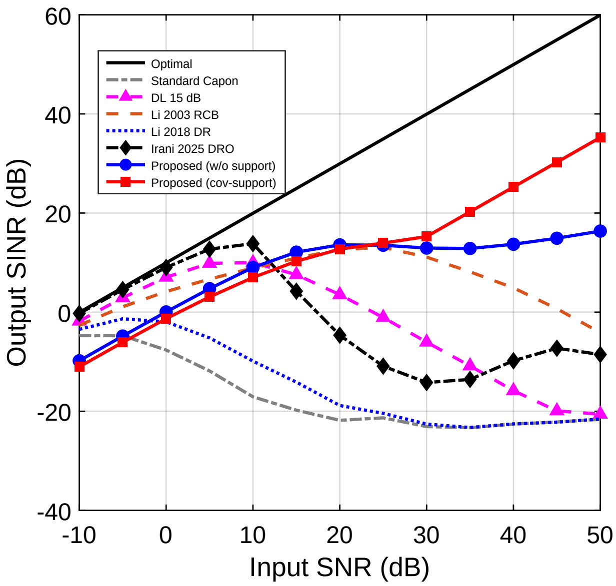

Fig. 2. Performance comparison under 4° angle mismatch.

The robust Capon beamformer (RCB) [11] performs well for mismatches up to 4°, after which its SINR drops rapidly—a direct consequence of the actual mismatch exceeding its fixed spherical uncertainty set ($\epsilon = 3.0$), which was designed for a 2° mismatch. The distributionally robust MVDR beamformer (DR-MVB) [14] degrades sharply beyond a 2° mismatch, falling well below diagonal loading. At moderate mismatch levels, its performance is limited by the fact that it only models steering vector uncertainty and neglects the covariance matrix contamination; at larger mismatches, the steering vector error itself exceeds the tolerance of its distributionally robust constraint, rendering the beamformer unable to maintain an effective array response in the true signal direction. The moment-based DRO beamformer (M-DRO) [15] remains nearly constant across the entire mismatch range; however, this is solely because its angular sector $[0^\circ, 6^\circ]$ is centered at the true DOA (3°)—an idealized configuration that completely insulates the moment estimates from the mismatch. When the sector is centered at the presumed direction, the performance of M-DRO degrades substantially. In contrast, both of the proposed Wasserstein-2 DRO variants maintain robust and stable performance across the full range of angle mismatches without relying on any prior knowledge of the true DOA. The data-driven bootstrap calibration automatically scales the regularization parameters in response to the increased mismatch, preventing the catastrophic performance loss observed in fixed-prior methods. At a moderate SNR of 20 dB, the support-free DRO and the covariance-support DRO exhibit comparable performance—at 6° mismatch the output SINR of both variants remains at approximately 10 dB, well above SCB, DL, RCB, and DR-MVB, all of which drop significantly. The support constraint is not yet decisive at this SNR level; its benefit becomes pronounced only at high SNR, as demonstrated in Fig. 1 and Fig. 2. This result highlights a defining advantage of the proposed framework: it generalizes robustly across widely varying mismatch levels without any manual re-tuning—a capability that conventional fixed-prior and moment-based approaches fundamentally lack.

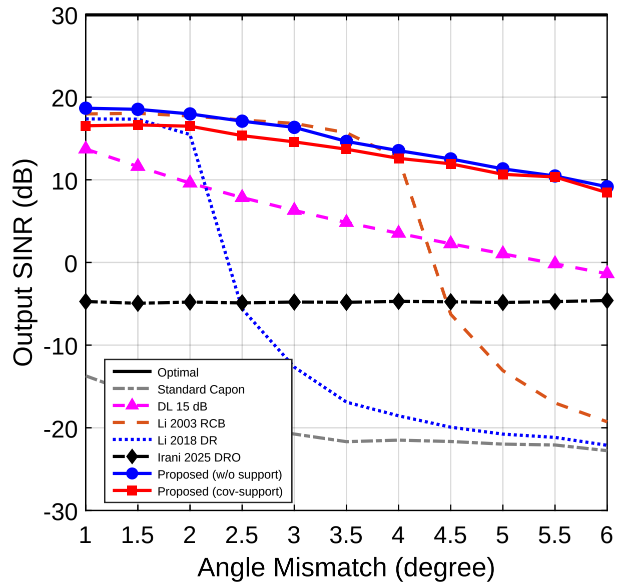

Fig. 3. Performance comparison versus angle mismatch at SNR = 20 dB.

Fig. 4 illustrates the output SINR versus the number of snapshots for the two proposed variants—the support-free DRO and the covariance-support DRO—under a 2° angle mismatch at a fixed input SNR of 10 dB. The number of snapshots is varied from 20 to 100. As described in Section IV, the bootstrap-based calibration procedure employs a sliding window of length $L = 3N = 30$ to preprocess the training data. When the number of available snapshots is less than L, the window length is reduced to half the number of snapshots to accommodate the limited sample size while preserving the statistical validity of the quantile estimation. Even with as few as 20 snapshots, both variants deliver an output SINR above 9 dB, and the performance improves steadily as the number of snapshots increases. By approximately 110 snapshots, the output SINR of both methods exceeds 14 dB, demonstrating that the bootstrap-based quantile calibration can reliably extract robust parameter estimates from limited training data. This steady improvement with increasing snapshots can be attributed to the fact that a larger sample size provides a more accurate empirical distribution of the sample covariance matrices, allowing the bootstrap procedure to produce a tighter and more reliable estimate of the Wasserstein radius ρ_1, which in turn reduces unnecessary conservatism in the regularization. Throughout the entire snapshot range, the two variants exhibit comparable performance at this moderate SNR level, consistent with the observations in Figs. 1–3 that the support constraint becomes decisive primarily at high SNR. This result confirms that the proposed data-driven calibration framework functions effectively under small-sample conditions, maintaining robust performance without requiring empirically tuned parameters.

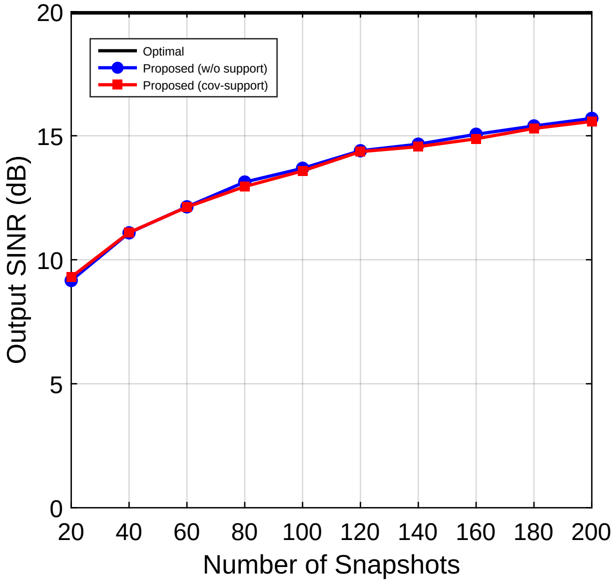

Fig. 4. Performance comparison versus number of snapshots under a 2° angle mismatch at SNR = 10 dB.

VI. Conclusion

This paper proposed a unified distributionally robust beamforming framework based on the Wasserstein-2 distance, which simultaneously models the distributional uncertainties in both the array steering vector and the interference-plus-noise covariance matrix. A physically motivated trace support constraint was introduced on the covariance ambiguity set, effectively capping the worst-case interference-plus-noise power at physically realizable levels and thereby overcoming the inherent limitation of non-reconstruction-based beamformers at high SNR—a deficiency that conventional diagonal loading and moment-based DRO approaches fail to address. A bootstrap-based quantile calibration strategy was developed to determine all Wasserstein radii and support-set parameters directly from training snapshots, providing rigorous finite-sample statistical guarantees and transitioning the selection of robustness parameters from empirical tuning to a principled, data-driven paradigm. Extensive simulations confirmed that the proposed method consistently outperforms existing robust beamformers across a wide range of mismatch conditions. The proposed framework establishes a principled, data-driven foundation for robust beamforming. Future research can naturally extend this foundation, for instance, by integrating it with reconstruction-based techniques or by establishing a theoretical link between the Wasserstein radius and generalization error to enable fully automatic confidence-level selection.

References

[1] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[2] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, 2010.

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

[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 Processing Magazine*, vol. 40, no. 4, pp. 118–131, 2023.

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

[6] J. Li and P. Stoica, *Robust adaptive beamforming*. Wiley Online Library, 2006.

[7] B. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.

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

[9] S.-J. Kim, A. Magnani, A. Mutapcic, S. P. Boyd, and Z.-Q. Luo, "Robust beamforming via worst-case sinr maximization," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1539–1547, 2008.

[10] R. Lorenz and S. Boyd, "Robust minimum variance beamforming," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1684–1696, 2005.

[11] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[12] X. Zhang, Y. Li, N. Ge, and J. Lu, "Robust minimum variance beamforming under distributional uncertainty," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2514–2518.

[13] X. Zhang, Q. Feng, N. Ge, and J. Lu, "Distributionally robust chance-constrained minimum variance beamforming," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2881–2885.

[14] B. Li, Y. Rong, J. Sun, and K. L. Teo, "A distributionally robust minimum variance beamformer design," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 105–109, 2018.

[15] K. H. Irani, Y. Huang, and S. A. Vorobyov, "Sinr maximizing distributionally robust adaptive beamforming," *IEEE Transactions on Signal Processing*, vol. 73, pp. 2542–2557, 2025.

[16] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1, pp. 115–166, 2018.

[17] J. Blanchet, Y. Kang, and K. Murthy, "Robust wasserstein profile inference and applications to machine learning," *Journal of Applied Probability*, vol. 56, no. 3, pp. 830–857, 2019.

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

[19] S. Wang, W. Dai, and G. Ye Li, "Distributionally robust adaptive beamforming," *IEEE Transactions on Signal Processing*, vol. 73, pp. 2981–2997, 2025.

[20] K. H. Irani, S. A. Vorobyov, and Y. Huang, "Wasserstein distributionally robust adaptive beamforming," 2025. [Online]. Available: https://arxiv.org/abs/2506.01154

[21] K. H. Irani, S. A. Vorobyov, and Y. Huang, "Learning-enhanced distributionally robust adaptive beamforming," in *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2026, pp. 20 726–20 730.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[23] A. Shapiro, "On duality theory of conic linear problems," in *Semi-Infinite Programming: Recent Advances*, ser. Nonconvex Optimization and Its Applications, M. Á. Goberna and M. A. López, Eds. Dordrecht: Kluwer Academic Publishers, 2001, no. 57, pp. 135–165.

[24] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[25] P. D. Tao and L. H. An, "Convex analysis approach to dc programming: theory, algorithms and applications," *Acta mathematica vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.