Distributionally Robust Optimization for STAP With Finite Samples

YALONG WANG, Graduate Student Member, IEEE
XUEJING ZHANG, Member, IEEE
ZHIHANG WANG, Member, IEEE
JUN LI
ZISHU HE, Senior Member, IEEE
University of Electronic Science and Technology of China , Chengdu, China

Drawing on the minimization of worst-case maximum likelihood (ML) estimation, this article develops a robust inverse clutter-plus-noise covariance matrix (CNCM) estimator for space–time adaptive processing against Gaussian clutter background at low sample support without any prior knowledge. Leveraging the nonconvex uncertainty set for CNCMs, we formulate a distributionally robust optimization-based ML estimation problem with the Wasserstein metric. We validate that the resulting nonconvex problem is algorithmically tractable. To achieve this, we reformulate the problem as a finite-dimensional semidefinite program. To pursue lower computational complexity, we establish a closed-form solution framework by imposing the rotation-equivariant property. We theoretically prove the existence and uniqueness of the solution and address the challenge of adaptively choosing the uncertainty set size. Importantly, the solution composes a nonlinear shrinkage estimator that inherently preserves the order of sample eigenvalues without additional operations. Experiments with both simulated and measured clutter data confirm the superiority of the proposed estimator in terms of estimation accuracy and robustness.

Received 28 October 2024; revised 7 March 2025; accepted 28 April 2025. Date of publication 2 May 2025; date of current version 13 October 2025.

DOI. No. 10.1109/TAES.2025.3566360

Refereeing of this contribution was handled by W. Liu.

This work was supported in part by the National Natural Science Foundation of China under Grant 62101101, Grant 62031007, Grant 62231006, and Grant 62371093, in part by Sichuan Science and Technology Program under Grant 2024NSFSC1433, and in part by Peng Cheng Shang Xue Education Fund under Grant XY2021602.

Authors' address: Yalong Wang, Xuejing Zhang, Zhihang Wang, Jun Li, and Zishu He are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, E-mail: (wang_yal@163.com; zhangxuejing@uestc.edu.cn; zhihang_w@uestc.edu.cn; lijunsc@uestc.edu.cn; zshe@uestc.edu.cn) *(Corresponding author: Xuejing Zhang.)*

I. INTRODUCTION

Space–time adaptive processing (STAP) is a powerful technology for moving platform radar systems, designed to suppress strong clutter and enhance weak target detection [1], [2], [3], [4], [5], [6]. Clutter-plus-noise covariance matrix (CNCM) estimation is crucial for implementing adaptive filtering in STAP. Generally, the sample covariance matrix (SCM) also well known as the maximum likelihood (ML) estimator for Gaussian clutter is employed. However, the Reed–Mallett–Brennan rule demonstrates its limitations, showing that the number of independently and identically distributed (IID) training samples should be at least twice the system degrees of freedom (DoFs) to ensure that the output signal-to-clutter-plus-noise ratio (SCNR) loss is less than 3 dB [7]. This poses significant challenges in actual clutter environments, especially when the number of available samples is further diminished due to the removal of outliers [8]. Consequently, developing robust estimators in uncertain and heterogeneous clutter environments with finite samples is a key research point for STAP.

To mitigate the dependence on the number of training samples, a series of suboptimal STAP methods has been proposed, which can be primarily categorized into three groups: reduced-dimension (RD) STAP [2], [9], [10], [11], [12], [13], [14], reduced-rank (RR) STAP [15], [16], [17], [18], and direct data domain (D3) STAP [19], [20]. RD-STAP and RR-STAP utilize well-designed transformation matrices to convert full-dimensional data into local-dimensional data or low-dimensional subspaces, respectively, so that the number of training samples is reduced to twice the local system DoFs or clutter rank. However, for conformal arrays or nonside-looking arrays, designing and selecting favorable auxiliary channels and accurately estimating the clutter rank remain challenges in practical applications. D3-STAP addresses the issue of limited samples by using only the data from the cell under test (CUT). However, this approach faces challenges, such as reduced system DoFs, sensitivity to channel errors, and the risk of target self-cancellation.

To improve CNCM estimation accuracy with limited samples, regularized ML estimators modify the SCM by incorporating prior information [21], [22], [23], [24], [25], [26], [27], [28]. Knowledge-aided STAP enhances detection performance by loading the prior CNCM onto the SCM [29], [30], [31], [32]. Naturally, the performance of these estimators is significantly influenced by the assumptions made about the structured model, the quality of prior knowledge, and the choice of the loading factor. Recently, shrinkage estimators have garnered significant attention due to their precise estimation performance and lower computational complexity. They can be classified into two categories: linear shrinkage estimators [33], [34], [35], [36], [37] and nonlinear shrinkage estimators [38], [39], [40], [41], [42], [43]. Linear shrinkage estimators aim to modify the sample eigenvalues by utilizing identical shrinkage density, while the nonlinear shrinkage estimators modify the sample eigenvalues by leveraging individualized

shrinkage density, thereby offering more opportunities to enhance the accuracy of CNCM estimation [39]. Besides, sparse recovery STAP and intelligent STAP can accurately reconstruct CNCM with very few samples, even with just one sample [38], [44], [45], [46], [47], [48], [49], [50], [51]. However, their application in practical radar systems is constrained by computational complexity and feasibility issues.

For uncertain and heterogeneous clutter environments, although the aforementioned methods improve the heterogeneous clutter suppression performance to some extent, their development often relies on prior information, such as prior CNCM, clutter rank, dictionary learning matrix, etc. However, in uncertain clutter environments, the reliability of such prior information is difficult to maintain, significantly compromising the robustness of these methods. Considering the powerful ability of distributionally robust optimization (DRO) theory in addressing uncertain optimization models [52], along with its successful applications in machine learning [53], automatic control [54], adaptive beamforming [55], and other fields, this article introduces the DRO theory into STAP and develops a precise and robust inverse CNCM estimator for Gaussian clutter background at low sample support without any prior information, named DRO-STAP.

Drawing inspiration from [56] and leveraging the advantages of the Wasserstein metric, such as tractability, finite sample support, and robust out-of-sample performance guarantees [54], we formulate a Wasserstein DRO problem that minimizes worst-case ML estimation within a nonconvex uncertainty set for CNCMs. To address the resulting nonconvex problem, we reformulate it into a finite-dimensional convex program and subsequently recast it into a tractable semidefinite program (SDP). To improve computational efficiency, we derive a closed-form solution framework using rotation equivariance, theoretically prove the existence and uniqueness of the solution, and address the challenge of choosing the uncertainty set size. The closed-form solution amounts to a nonlinear shrinkage estimator, which inherently preserves the order of sample eigenvalues without additional operations. Finally, we conduct a series of numerical experiments using both simulated and measured clutter data to validate the effectiveness and robustness of DRO-STAP.

The rest of this article is organized as follows. Section II denotes the signal model, STAP criterion, and problem formulation. Section III derives the DRO-based ML estimation framework, recasts it into a tractable form, presents a closed-form solution, discusses the choice of uncertainty set size, and examines the computational complexity. Section IV provides the numerical experiments to verify the effectiveness and robustness of DRO-STAP. Finally, Section V concludes this article.

Notations: In this article, we use lightface for scalar a, lower case boldface for vector $\mathbf{a}$, and upper case for matrix $\mathbf{A}$. $\mathrm{Tr}(\mathbf{A})$ and $\mathbf{A}^{-1}$ denotes the trace and inverse of matrix $\mathbf{A}$, respectively. $\mathrm{diag}(\mathbf{a})$ is the diagonalization of the vector $\mathbf{a}$. $\mathbf{a} \sim \mathcal{CN}(\mathbf{0}, \mathbf{A})$ means that the random vector $\mathbf{a}$ follows a circularly symmetric complex Gaussian distribution with zero-mean and covariance $\mathbf{A}$. $\mathbf{a}^{\downarrow}$ and $\mathbf{a}^{\uparrow}$ indicate that the elements in $\mathbf{a}$ are arranged in descending and ascending order, respectively. Double line body $\mathbb{C}$ means the complex numbers domain. $(\cdot)^*, (\cdot)^T, (\cdot)^H, \otimes$, and $\mathbb{E}\{\cdot\}$ denote the conjugate, transpose, conjugate transpose, Kronecker product, and statistical expectation operators, respectively. $\mathbf{I}_N$ is the N-dimensional identity matrix. $|\cdot|$ represents the absolute value of a scalar or the determinant of a matrix. $\|\cdot\|_2$ is the l_2 norm. $\mathbb{H}_{+}^{N}$ ($\mathbb{H}_{++}^{N}$) denotes the cone of N-dimensional positive semidefinite (definite) Hermitian matrix. $\mathbf{A} \succ \mathbf{B}$ ($\mathbf{A} \succeq \mathbf{B}$) indicates that $\mathbf{A} - \mathbf{B}$ is positive semidefinite (definite). $\Pr\{\cdot\}$ represents the probability of an event. $a \xrightarrow{\mathrm{p}} b$ denotes that a converges to b in probability. $\propto$ means that the quantities on its both sides are directly proportional.

II. PRELIMINARIES

A. Signal Model and STAP Principle

For the sake of illustration, we consider a side-looking airborne radar equipped with N-element uniform linear array, the interspacing is $d = \lambda/2$, and λ represents the radar wavelength. The platform flies horizontally at height H with constant velocity v_p. During a coherent processing interval, M pulses with identical pulse repetition frequency f_r are emitted to detect targets in areas of interest. After preprocessing, such as down-conversion, analog-to-digital conversion, and matched filtering (MF), the received baseband snapshot data $\mathbf{z} \in \mathbb{C}^{MN}$ for a given range cell can be mathematically formulated as

$$\mathbf{z} = \alpha_t \mathbf{s}_t + \mathbf{z}_c + \mathbf{z}_n \tag{1}$$

where $\alpha_t \mathbf{s}_t$ denotes the potential target, α_t is the unknown complex amplitude, and $\mathbf{s}_t$ denotes its space–time steering vector. $\mathbf{z}_c$ represents the signal-dependent clutter echo, without considering the range ambiguity, and it can be formulated via the classical discrete integral clutter model [1]

$$\mathbf{z}_c = \sum_{j=1}^{N_c} \alpha_j \mathbf{s}_{c,j} = \sum_{j=1}^{N_c} \alpha_j \mathbf{a}_s(f_{c,j}^s) \otimes \mathbf{a}_d(f_{c,j}^d) \tag{2}$$

where N_c is the number of IID clutter patches. α_j and $\mathbf{s}_{c,j}$ are complex amplitude and space–time steering vector of the jth clutter patch, respectively. $\mathbf{a}_s(f_{c,j}^s)$ and $\mathbf{a}_d(f_{c,j}^d)$ denote the spatial and Doppler steering vectors of the corresponding clutter patch, respectively. Having

$$\mathbf{a}_s(f_{c,j}^s) = \frac{1}{\sqrt{N}}[1, e^{j2\pi f_{c,j}^s}, \ldots, e^{j2\pi(N-1)f_{c,j}^s}]^T \tag{3a}$$

$$\mathbf{a}_d(f_{c,j}^d) = \frac{1}{\sqrt{M}}[1, e^{j2\pi f_{c,j}^d}, \ldots, e^{j2\pi(M-1)f_{c,j}^d}]^T \tag{3b}$$

$f_{c,j}^s = d \sin\theta_j \sin\varphi_j/\lambda$ and $f_{c,j}^d = 2v_p \sin\theta_j \sin\varphi_j/\lambda f_r$, respectively, denote the normalized spatial frequency and normalized Doppler frequency of the jth clutter patch. θ_j and φ_j are the elevation and azimuth angles of the corresponding clutter patch. $\mathbf{z}_n$ represents the zero-mean complex Gaussian white noise with variance matrix $\delta_n^2 \mathbf{I}_{MN}$, and δ_n^2 is the noise

power. The ideal CNCM $\mathbf{R}$ can thus be computed by

$$\begin{aligned}\mathbf{R} &= \mathbb{E}\{(\mathbf{z}_c + \mathbf{z}_n)(\mathbf{z}_c + \mathbf{z}_n)^H\} \\ &= \sum_{j=1}^{N_c} \mathbb{E}\{|\alpha_j|^2\}\mathbf{s}_{c,j}\mathbf{s}_{c,j}^H + \delta_n^2 \mathbf{I}_{MN}. \end{aligned} \tag{4}$$

By maximizing the output SCNR, the optimal STAP filter weight vector can be derived through the minimum variance distortionless response rule, giving

$$\mathbf{w} = \eta \mathbf{R}^{-1}\mathbf{s}_t \tag{5}$$

where $\eta = 1/\mathbf{s}_t^H \mathbf{R}^{-1}\mathbf{s}_t$ represents the normalized factor.

B. Problem Formulation and Analysis

Note that achieving the optimal STAP filter in (5) is extremely challenging, as $\mathbf{R}$ is unknown in practice. The widely adopted approach is estimating $\hat{\mathbf{R}}$ through L IID target-free training samples $\{\mathbf{z}_l\}_{l=1}^L$. This estimate $\hat{\mathbf{R}}$ is then used to implement a suboptimal STAP filter $\hat{\mathbf{w}}$ for target detection. The detection probability is monotonic related to the estimation accuracy of $\hat{\mathbf{R}}$ [7]. For STAP applications, $\{\mathbf{z}_l\}_{l=1}^L$ is generally modeled as a zero-mean circularly symmetric complex Gaussian distribution [41], i.e., $\mathbf{z}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$. The SCM also well known as the unconstrained ML estimation can be denoted as

$$\hat{\mathbf{R}}_{\text{SCM}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{z}_l \mathbf{z}_l^H. \tag{6}$$

However, in scenarios with low samples ($L < 2MN$ or even $L < MN$), fixed-dimensional asymptotics become inapplicable, resulting in significant performance degradation with $\hat{\mathbf{R}}_{\text{SCM}}$. Thereby, to improve covariance estimation performance with finite samples, some prior knowledge, including noise power, clutter rank, and covariance structure, is often incorporated into existing estimators [41], [42], [43]. However, these methods may lack robustness when there are errors in the prior information or the assumed covariance model does not hold anymore. Moreover, existing methods typically derive the STAP filter through the following steps: covariance estimation, inversion of the covariance matrix, and filter design. In contrast, directly estimating the inverse covariance matrix with limited samples may be more effective for enhancing STAP performance.

III. DRO-BASED INVERSE CNCM ML ESTIMATION FOR STAP

In this section, we develop a precise inverse CNCM ML estimation method for STAP at low sample support. It is derived from a DRO-based ML estimation model, which we call DRO-STAP. Particularly, DRO-STAP does not rely on any prior knowledge and covariance structure, making it more robust in clutter suppression. To pursue computational efficiency, we also present the closed-form solution for DRO-STAP, and it appertains a nonlinear shrinkage estimator. Finally, the choice of uncertainty set size and the computational complexity are discussed, guiding the practical application of DRO-STAP.

A. Inverse CNCM ML Estimation

With the zero-mean complex circular Gaussian clutter assumption, the joint probability density function (PDF) of L IID samples $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_L]$ is given by

$$\begin{aligned} p(\mathbf{Z}; \mathbf{R}) &= \frac{1}{\pi^{MNL}|\mathbf{R}|^L} \exp\left(-\sum_{l=1}^{L} \mathbf{z}_l^H \mathbf{R}^{-1}\mathbf{z}_l\right) \\ &\propto |\mathbf{R}|^{-L} \exp\left(-L\text{Tr}(\mathbf{R}^{-1}\hat{\mathbf{R}}_{\text{SCM}})\right). \end{aligned} \tag{7}$$

To estimate $\mathbf{R}^{-1}$, for the sake of clarity, we first let $\mathbf{X} = \mathbf{R}^{-1}$, and the joint PDF $p(\mathbf{Z}; \mathbf{R})$ evolves to

$$p(\mathbf{Z}; \mathbf{R}) \triangleq p(\mathbf{Z}; \mathbf{X}) \propto |\mathbf{X}|^L \exp\left(-L\text{Tr}(\mathbf{X}\hat{\mathbf{R}}_{\text{SCM}})\right). \tag{8}$$

Maximizing the logarithm of the likelihood function $p(\mathbf{Z}; \mathbf{X})$, one can access the ML estimation of inverse CNCM $\mathbf{X}$. This is equivalent to solving the following unconstrained convex optimization problem:

$$\hat{\mathbf{X}}_{\text{ML}} = \arg\min_{\mathbf{X}\succ\mathbf{0}} \ \text{Tr}(\mathbf{X}\hat{\mathbf{R}}_{\text{SCM}}) - \ln|\mathbf{X}|. \tag{9}$$

If $L \geq MN$, that is to say, $\hat{\mathbf{R}}_{\text{SCM}}$ is nonsingular. Then, (9) can be minimized in a closed-form solver $\hat{\mathbf{X}}_{\text{ML}} = \hat{\mathbf{R}}_{\text{SCM}}^{-1}$. However, if $L < MN$, $\hat{\mathbf{R}}_{\text{SCM}}$ is singular and (9) becomes unbounded.

To solve problem (9) efficiently at $L < MN$, various regularized (constrained) ML estimation algorithms have been developed [21], [22], [23], [24], [25], [26], [27], [28]. Among these methods, the core idea is also to incorporate prior information, such as noise power, clutter rank, and the prior covariance matrix, to regulate problem (9). Indeed, these approaches enhance the convergence performance of STAP. Similarly, the robustness of these methods has deteriorated.

B. Problem Formulation for DRO-Based Inverse CNCM ML Estimation

In contrast to the aforementioned ML estimation algorithms, our approach does not incorporate any prior information about clutter and noise. Instead, we reformulate problem (9) based on DRO theory to enhance the robustness of the estimator. Given that the true distribution of the sample $\mathbf{z} \in \{\mathbf{z}_l\}_{l=1}^L$ is $\mathbb{T} = \mathcal{CN}(\mathbf{0}, \mathbf{R})$, the optimal estimate of $\mathbf{X} = \mathbf{R}^{-1}$ can be obtained by solving the following optimization problem:

$$\hat{\mathbf{X}}_{\text{opt}} = \arg\min_{\mathbf{X}\succ\mathbf{0}} \ \mathbb{E}_{(\mathbb{T})}\left\{\mathbf{z}^H\mathbf{X}\mathbf{z}\right\} - \ln|\mathbf{X}|. \tag{10}$$

Optimization model (10) represents the conditional expectation ML estimation under true distribution $\mathbb{T}$. Since true distribution $\mathbb{T}$ is unknown, we cannot access the optimal estimate $\hat{\mathbf{X}}_{\text{opt}}$ by minimizing (10). It is important to note that the ML estimation in (9) is equivalent to replacing $\mathbb{T}$ with the empirical distribution $\hat{\mathbb{T}} = \mathcal{CN}(\mathbf{0}, \hat{\mathbf{R}}_{\text{SCM}})$, thus simplifying problem (10). To achieve a precise and robust

inverse CNCM estimator, we aim to constrain the statistical distance between the empirical distribution $\tilde{\mathbb{T}}$ and true distribution $\mathbb{T}$.

REMARK 1 It is known that various statistical distances, for example, Kullback–Leibler divergence, relative entropy, and the Wasserstein metric, can quantify the cost of transforming one distribution into another. Among these statistical distances, the Wasserstein distance has garnered significant attention due to numerous advantages, such as tractability, finite sample support, and robust out-of-sample performance guarantees [54].

Drawing on the analysis from Remark 1, we take the Wasserstein distance into account for heterogeneous and uncertain clutter environments, defined as follows.

DEFINITION 1 (WASSERSTEIN DISTANCE) The Wasserstein distance of order p between any two distributions $\mathbf{z}_1 \sim \mathbb{T}_1$ and $\mathbf{z}_2 \sim \mathbb{T}_2$ supported on $\mathcal{Z} \subseteq \mathbb{C}^N$ is defined as follows:

$$W_p(\mathbb{T}_1, \mathbb{T}_2) = \min_{\tau \in \mathcal{T}(\mathbb{T}_1, \mathbb{T}_2)} \left\{ \left(\int_{\mathcal{Z}\times\mathcal{Z}} \|\mathbf{z}_1 - \mathbf{z}_2\|^p d\tau(\mathbf{z}_1, \mathbf{z}_2) \right)^{\frac{1}{p}} \right\} \quad (11)$$

where $\mathcal{T}(\mathbb{T}_1, \mathbb{T}_2)$ denotes the joint distribution of $\mathbf{z}_1$ and $\mathbf{z}_2$ with marginals $\mathbb{T}_1$ and $\mathbb{T}_2$, respectively. $\|\cdot\|$ could be an arbitrary norm defined on $\mathcal{Z}$.

By incorporating the Wasserstein distance constraint between the empirical distribution $\tilde{\mathbb{T}}$ and the true distribution $\mathbb{T}$, the modified ML estimation problem (10) can be re-expressed as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}\succ\mathbf{0}} \ \mathbb{E}_{(\tilde{\mathbb{T}})}\left\{\mathbf{z}^H\mathbf{X}\mathbf{z}\right\} - \ln|\mathbf{X}| \quad \text{s.t.} \ \ W_p(\tilde{\mathbb{T}}, \mathbb{T}) \le \tilde{\rho}. \quad (12)$$

Optimization model (12) represents the conditional expectation ML estimation under the empirical distribution $\tilde{\mathbb{T}}$ within the Wasserstein distance constraint, and $\tilde{\rho}$ denotes the Wasserstein distance control factor. Especially, when $\tilde{\rho} = 0$, problem (12) reduces to problem (10). Likewise, directly optimizing problem (12) remains infeasible since $\mathbb{T}$ is unknown, rendering $W_p(\tilde{\mathbb{T}}, \mathbb{T})$ uncomputable.

To overcome this challenge, following the approach in [56], let us first assume that the true distribution $\mathbb{T}$ is included in a distribution set $\mathbb{D}$ of all reference complex Gaussian distribution with different CNCMs, that is

$$\mathbb{T} \subseteq \mathbb{D} \in \mathcal{CN}_{\mathbf{0}}^{MN} \quad (13)$$

where $\mathcal{CN}_{\mathbf{0}}^{MN}$ denotes the family of all MN-dimensional zero-mean complex circular Gaussian distribution. Based on a given empirical distribution $\tilde{\mathbb{T}}$ and an appropriate uncertainty set size ρ, we can construct the distribution set $\mathbb{D}$ as follows:

$$\mathcal{S} = \left\{ \mathbb{D} \in \mathcal{CN}_{\mathbf{0}}^{MN} \left| \begin{array}{l} \Pr\{\mathbb{T} \subseteq \mathbb{D}\} = 1 \\ \mathbb{E}_{(\mathbb{D})}\{\mathbf{R}\} \succeq \mathbf{0} \\ W_p(\tilde{\mathbb{T}}, \mathbb{D}) \le \rho \end{array} \right. \right\}. \quad (14)$$

Since the empirical distribution $\tilde{\mathbb{T}}$ is close to the true distribution $\mathbb{T}$, we can ensure that the true distribution $\mathbb{T}$ is always contained within the well-constructed distribution set $\mathbb{D}$, that is, $\Pr\{\mathbb{T} \subseteq \mathbb{D}\} = 1$. At this point, we can only guarantee the inclusion of $\mathbb{T}$ in $\mathbb{D}$ but do not know its exact form. To address this uncertainty, we can achieve a robust inverse CNCM estimation by minimizing the worst-case (maximum) ML estimation within the distribution set $\mathbb{D}$, thereby circumventing the challenge of the unknown $\mathbb{T}$. Ultimately, given L samples, it is natural to formulate a DRO-based ML estimation problem for the inverse CNCM as follows:

$$\min_{\mathbf{X}} \ \max_{\mathbb{D}} \ \mathbb{E}_{(\mathbb{D})}\left\{\mathrm{Tr}(\mathbf{X}\mathbf{z}\mathbf{z}^H)\right\} - \ln|\mathbf{X}| \quad \text{s.t.} \ \ \mathbb{D} \in \mathcal{S}, \mathbf{X} \succ \mathbf{0}. \quad (15)$$

REMARK 2 With the Gaussian clutter assumption, optimization model (15) minimizes the maximum (worst-case) conditional expectation ML estimation within the uncertainty set $\mathbb{D}$ to achieve a precise and robust estimator without any prior information. It is important to note that optimization model (15) can also be further extended to non-Gaussian clutter backgrounds. In short, we can model heavy-tailed non-Gaussian clutter by the complex elliptically symmetric distributions [57], use normalized SCM [58] or Tyler's estimator [59] to construct the empirical distribution $\tilde{\mathbb{T}}$, and then construct the minimax conditional expectation ML estimation model similar to (15) within the corresponding uncertainty set $\tilde{\mathbb{D}}$. The problem of estimating the inverse CNCM under non-Gaussian clutter remains further investigated but is beyond the scope of this article.

Clearly, the DRO problem (15) is challenging to solve for two primary reasons. First, the distributional uncertainty set constraint $\mathbb{D} \in \mathcal{S}$ is nonconvex, and directly evaluating the Wasserstein distance between two arbitrary distributions is often impractical, especially with limited samples [60]. Second, problem (15) involves an infinite-dimensional minimax optimization, which is NP-hard. To overcome these issues, the following section reveals that the DRO problem (15) is tractable and can be reformulated as a finite-dimensional SDP problem by leveraging the duality theory.

C. Convex Reformulation for DRO-Based Inverse CNCM ML Estimation

To reformulate the DRO problem (15), the central steps involve approximating the Wasserstein distance metric, re-expressing, and solving the inner maximum (worst-case) conditional expectation optimization problem. First, we propose employing the Gelbrich bound as a practical alternative for finite samples. The Gelbrich distance provides a lower bound for the type-2 ($p = 2$) Wasserstein distance with the Euclidean norm and applies to any nominal distribution. It is defined as follows.

DEFINITION 2 (GELBRICH BOUND [61]) For any two nominal distributions $\mathbb{T}_1$ and $\mathbb{T}_2$ with mean vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{C}^N$ and covariance matrices $\mathbf{R}_1, \mathbf{R}_2 \in \mathbb{H}_+^N$, the Gelbrich

distance is defined as follows:

$$G(\mathbb{T}_1, \mathbb{T}_2) = \sqrt{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 + B_W^2(\mathbf{R}_1, \mathbf{R}_2)} \leq W_2(\mathbb{T}_1, \mathbb{T}_2) \tag{16}$$

where

$$B_W^2(\mathbf{R}_1, \mathbf{R}_2) = \mathrm{Tr}\left(\mathbf{R}_1 + \mathbf{R}_2 - 2\left(\mathbf{R}_2^{\frac{1}{2}}\mathbf{R}_1\mathbf{R}_2^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \tag{17}$$

represents the squared Bures–Wasserstein distance. Furthermore, the inequality in (16) holds with equality if $\mathbb{T}_1$ and $\mathbb{T}_2$ are elliptical with the same density generator. It is evident that $G(\mathbb{T}_1, \mathbb{T}_2)$ is real and remains bounded, irrespective of the deterioration of $\mathbb{T}_1$ or $\mathbb{T}_2$ (i.e., $\mathbf{R}_1$ or $\mathbf{R}_2$ is rank deficient).

Herein, we consider zero-mean clutter, that is, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, and then the Gelbrich distance simplifies to the Bures–Wasserstein distance (hereinafter referred to as Wasserstein distance). Leveraging the empirical distribution $\hat{\mathbb{T}} = \mathcal{CN}(\mathbf{0}, \hat{\mathbf{R}}_{\mathrm{SCM}})$, the distribution set $\mathbb{D}$ is redefined as follows:

$$\tilde{\mathcal{S}} = \left\{ \mathbb{D} \in \mathcal{CN}_{\mathbf{0}}^{MN} \middle| \begin{array}{l} \Pr\{\mathbb{T} \subseteq \mathbb{D}\} = 1 \\ \mathbb{E}_{(\mathbb{D})}\{\mathbf{R}\} \succeq \mathbf{0} \\ B_W(\hat{\mathbf{R}}_{\mathrm{SCM}}, \mathbb{E}_{(\mathbb{D})}\{\mathbf{R}\}) \leq \rho \end{array} \right\}. \tag{18}$$

To proceed, disregarding irrelevant variables, the inner maximization expectation problem can be expressed as

$$\max_{\mathbb{D}} \ \mathbb{E}_{(\mathbb{D})}\left\{\mathrm{Tr}(\mathbf{X}\mathbf{z}\mathbf{z}^H)\right\} \quad \text{s.t.} \ \ \mathbb{D} \in \tilde{\mathcal{S}}. \tag{19}$$

By introducing an auxiliary variable $\mathbf{Y} \succeq \mathbf{0}$, we find that solving problem (19) is equivalent to solving the following optimization problem:

$$\begin{aligned} \max_{\mathbf{Y}} \ & \mathrm{Tr}(\mathbf{X}\mathbf{Y}) \\ \text{s.t.} \ & \mathbf{Y} \succeq \mathbf{0}, B_W(\hat{\mathbf{R}}_{\mathrm{SCM}}, \mathbf{Y}) \leq \rho. \end{aligned} \tag{20}$$

By adding a penalty parameter $\varepsilon > 0$ for the Wasserstein distance constraint, we can formulate the Lagrangian problem for (20) as follows:

$$\begin{aligned} \max_{\mathbf{Y}} \ \min_{\varepsilon} \ & \mathcal{L}(\mathbf{Y}, \varepsilon) \\ \text{s.t.} \ & \mathbf{Y} \succeq \mathbf{0}, \varepsilon > 0. \end{aligned} \tag{21}$$

Using the equality in (17), the Lagrangian function $\mathcal{L}(\mathbf{Y}, \varepsilon)$ is denoted as

$$\begin{aligned} \mathcal{L}(\mathbf{Y}, \varepsilon) =& \ \mathrm{Tr}(\mathbf{X}\mathbf{Y}) + \varepsilon(\rho^2 - B_W^2(\hat{\mathbf{R}}_{\mathrm{SCM}}, \mathbf{Y})) \\ =& \ \mathrm{Tr}(\mathbf{Y}(\mathbf{X} - \varepsilon\mathbf{I}_{MN})) + \varepsilon(\rho^2 - \mathrm{Tr}(\hat{\mathbf{R}}_{\mathrm{SCM}})) \\ & + 2\varepsilon\mathrm{Tr}\left(\left(\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\mathbf{Y}\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right). \end{aligned} \tag{22}$$

Further, the dual problem for (21) is cast as

$$\begin{aligned} \min_{\varepsilon} \ \max_{\mathbf{Y}} \ & \mathcal{L}_{\mathbf{Y},\varepsilon} + \mathcal{L}_{\varepsilon} \\ \text{s.t.} \ & \mathbf{Y} \succeq \mathbf{0}, \varepsilon > 0. \end{aligned} \tag{23}$$

For notational simplicity, let $\mathbf{X}_\varepsilon = \mathbf{X} - \varepsilon\mathbf{I}_{MN}$, where

$$\mathcal{L}_{\mathbf{Y},\varepsilon} = \mathrm{Tr}(\mathbf{Y}\mathbf{X}_\varepsilon) + 2\varepsilon\mathrm{Tr}\left(\left(\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\mathbf{Y}\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \tag{24a}$$

$$\mathcal{L}_{\varepsilon} = \varepsilon(\rho^2 - \mathrm{Tr}(\hat{\mathbf{R}}_{\mathrm{SCM}})). \tag{24b}$$

The strong duality between (21) and (23) holds for any $\rho > 0$ in terms of the result for moment problems [56], [62]. At this point, the following proposition is executed to establish the basis for reformulating the convex form of the DRO problem (15).

PROPOSITION 1 When $\mathbf{X}_\varepsilon \prec \mathbf{0}$, the inner maximization subproblem in (23), i.e.,

$$\max_{\mathbf{Y}} \ \mathcal{L}_{\mathbf{Y},\varepsilon} \quad \text{s.t.} \ \ \mathbf{Y} \succeq \mathbf{0} \tag{25}$$

can be re-expressed as a convex problem, its optimal solution of $\mathbf{Y}$ can be obtained via the first-order optimality condition given by

$$\mathbf{Y}^\star = \varepsilon^2\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}}\mathbf{X}_\varepsilon^{-1} \tag{26}$$

and the corresponding maximum of (25) is equivalent to

$$\mathcal{L}_{\mathbf{Y}^\star,\varepsilon} = -\varepsilon^2\mathrm{Tr}(\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}}). \tag{27}$$

PROOF See Appendix A. ■

Ultimately, the infinite-dimensional problem (15) can be converted into the following finite-dimensional convex optimization problem via proposition 1, that is:

$$\begin{aligned} \min_{\mathbf{X},\varepsilon} \ & \mathcal{L}_{\varepsilon} - \varepsilon^2\mathrm{Tr}(\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}}) - \ln|\mathbf{X}| \\ \text{s.t.} \ & \mathbf{X} \succ \mathbf{0}, \mathbf{X}_\varepsilon \prec \mathbf{0}, \varepsilon > 0. \end{aligned} \tag{28}$$

Since the objective function of problem (28) includes inverse matrix $\mathbf{X}_\varepsilon^{-1}$, the nonlinear term $\varepsilon^2\mathrm{Tr}(\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}})$ is difficult to be solved directly. To do so, we introduce another auxiliary matrix variable $\mathbf{Q}$ and ensure that $\mathrm{Tr}(\mathbf{Q}) \succeq \varepsilon^2\mathrm{Tr}(-\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}})$, i.e., $\mathbf{Q} \succeq -\varepsilon\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\varepsilon$. By Schur's complement, problem (28) can be recast as the following SDP problem:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{Q},\varepsilon} \ & \mathcal{L}_{\varepsilon} + \mathrm{Tr}(\mathbf{Q}) - \ln|\mathbf{X}| \\ \text{s.t.} \ & \begin{bmatrix} -\mathbf{X}_\varepsilon & \varepsilon\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}} \\ \varepsilon\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}} & \mathbf{Q} \end{bmatrix} \succeq \mathbf{0} \\ & \mathbf{X} \succ \mathbf{0}, \mathbf{X}_\varepsilon \prec \mathbf{0}, \mathbf{Q} \succeq \mathbf{0}, \varepsilon > 0. \end{aligned} \tag{29}$$

Similarly, the SDP problem (29) can be solved by off-the-shelf solvers, such as SDPT3, SeDuMi, and MOSEK. Specifically, this process involves solving for $\mathbf{X}$, $\mathbf{Q}$, and ε, resulting in a total of $MN(MN+1)+1$ optimization variables. The computational complexity of solving SDP (29) by the above interior point solvers is $\mathcal{O}(M^6N^6)$ per iteration. Therefore, the solution of (29) obtained through the above solvers cannot be applied in real time, especially for large MN. Alternating direction multiplier method-based approach [63] may provide a lower computational complexity of solving problem (29), but its convergence and the choice of regularization parameter remain open challenges [64]. To pursue a lower computational complexity and a more stable solution, for any preset $\rho > 0$, the closed-form solution framework for penalty parameter ε

and inverse covariance matrix $\mathbf{X}$ will be introduced in the following section.

D. Closed-Form Solution for DRO-Based Inverse CNCM ML Estimation

Initially, note that without prior knowledge about the orientation of eigenvectors of $\mathbf{X}$, it is challenging to derive a closed-form solution for $\mathbf{X}$ directly. Fortunately, if $\hat{\mathbf{X}}$ belongs to the rotation-equivariant estimator, optimizing (28) in closed form over this class becomes possible. The rotation-equivariant estimator is defined as follows.

DEFINITION 3 (ROTATION-EQUIVARIANT ESTIMATOR [40]) For any unitary matrix, also referred to as a rotation matrix, $\mathbf{V} \in \mathbb{C}^{MN\times MN}$, $\hat{\mathbf{A}}$ is termed a rotation-equivariant estimator if $\hat{\mathbf{A}}(\mathbf{V}\hat{\mathbf{R}}_{\text{SCM}}\mathbf{V}^H) = \mathbf{V}\hat{\mathbf{A}}(\hat{\mathbf{R}}_{\text{SCM}})\mathbf{V}^H$.

This definition reveals that rotation-equivariant estimators share identical eigenvectors with SCM. Apparently, shrinkage estimators are rotation-equivariant. As discussed in Section I, unless otherwise stated, this article focuses on the research of nonlinear shrinkage estimators. These estimators begin with performing eigendecomposition on $\hat{\mathbf{R}}_{\text{SCM}}$

$$\hat{\mathbf{R}}_{\text{SCM}} = \mathbf{U}\text{diag}(\kappa_1, \kappa_2, \ldots, \kappa_{\text{MN}})\mathbf{U}^H \tag{30}$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{MN}]$ represents the eigenvector matrix of $\hat{\mathbf{R}}_{\text{SCM}}$, and $\boldsymbol{\kappa} = [\kappa_1, \kappa_2, \ldots, \kappa_{MN}]^T$, where $\kappa_1 \le \kappa_2 \le \cdots \le \kappa_{MN}$ denotes the corresponding eigenvalue vector.

As $L, MN \to \infty$ together, leveraging a proxy (i.e., Frobenius loss metric) to simplify the SCNR metric, the optimal shrinkage eigenvalues are derived by [40], [41]

$$\kappa_i^\star = \mathbf{u}_i^H\mathbf{R}\mathbf{u}_i,\ i = 1, 2, \ldots, MN \tag{31}$$

and the optimal (or Oracle) shrinkage estimator is

$$\hat{\mathbf{R}}_{\text{Ora}} = \mathbf{U}\text{diag}(\kappa_1^\star, \kappa_2^\star, \ldots, \kappa_{MN}^\star)\mathbf{U}^H. \tag{32}$$

Likewise, $\hat{\mathbf{R}}_{\text{Ora}}$ is also unavailable because $\mathbf{R}$ is unknown. Thus, several suboptimal individualized shrinkage schemes to obtain the shrinkage eigenvalues $\tilde{\boldsymbol{\kappa}}$ from $\boldsymbol{\kappa}$ are designed [38], [39], [40], [41], [43], they all can be formulated as follows:

$$\tilde{\boldsymbol{\kappa}} = \mathcal{F}_m(\boldsymbol{\kappa}) = [\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_{MN}]^T \tag{33}$$

where $\mathcal{F}_m(\cdot)$ is the mapping function of the corresponding shrinkage scheme. Finally, shrinkage estimators construct the covariance matrix through $\mathbf{U}$ and $\tilde{\boldsymbol{\kappa}}$

$$\hat{\mathbf{R}}_{\text{shr}} = \mathbf{U}\text{diag}(\tilde{\boldsymbol{\kappa}})\mathbf{U}^H. \tag{34}$$

This class of shrinkage estimators is exactly what we are interested in this article for deriving the closed-form solution of DRO-STAP, and inverse CNCM estimation $\hat{\mathbf{X}}$ can thus be written as

$$\hat{\mathbf{X}} = \mathbf{U}\text{diag}(\hat{\kappa}_{X,1}, \hat{\kappa}_{X,2}, \ldots, \hat{\kappa}_{X,MN})\mathbf{U}^H \tag{35}$$

where $\hat{\kappa}_{X,i}$, $i = 1, 2, \ldots, MN$, are the eigenvalues to be solved of $\hat{\mathbf{X}}$.

Subsequently, the connection (i.e., $\mathcal{F}_m(\cdot)$) between the DRO-based ML estimator and the shrinkage estimator is established, and $\hat{\kappa}_{X,i}$ is then attained effectively. Recall the re-expressed convex form of the DRO-based ML estimation problem (28), let

$$\begin{aligned}\mathcal{F}(\mathbf{X}, \varepsilon) = {} & \varepsilon(\rho^2 - \text{Tr}(\hat{\mathbf{R}}_{\text{SCM}})) - \ln|\mathbf{X}| \\ & - \varepsilon^2\text{Tr}((\mathbf{X} - \varepsilon\mathbf{I}_{MN})^{-1}\hat{\mathbf{R}}_{\text{SCM}})\end{aligned} \tag{36}$$

represents its objective function. Leveraging the eigenvector matrix $\mathbf{U}$ [as shown in (30)] shared with $\hat{\mathbf{R}}_{\text{SCM}}$, the optimal solution of $\mathbf{X}$ can be calculated by the first-order optimality condition

$$\begin{aligned}& \partial\mathcal{F}(\mathbf{X}, \varepsilon)/\partial\mathbf{X} \\ = {} & \varepsilon^2(\mathbf{X}^* - \varepsilon\mathbf{I}_{MN})^{-1}\hat{\mathbf{R}}_{\text{SCM}}^*(\mathbf{X}^* - \varepsilon\mathbf{I}_{MN})^{-1} - (\mathbf{X}^*)^{-1} \\ = {} & \sum_{i=1}^{MN}\frac{\varepsilon^2\kappa_i\kappa_{X,i} - (\kappa_{X,i} - \varepsilon)^2}{(\kappa_{X,i} - \varepsilon)^2\kappa_{X,i}}\mathbf{u}_i^*\mathbf{u}_i^T \\ = {} & \mathbf{0}\end{aligned} \tag{37}$$

and the optimal solution of $\kappa_{X,i}$ satisfies the following relationship:

$$\varepsilon^2\kappa_i\kappa_{X,i} = (\kappa_{X,i} - \varepsilon)^2\ ,\ i = 1, 2, \ldots, MN. \tag{38}$$

Obviously, we need to determine the optimal penalty parameter ε. Similar to (37), by utilizing the equality (38), the optimal solution of ε can be attained by the first-order optimality condition

$$\begin{aligned}& \partial\mathcal{F}(\mathbf{X}, \varepsilon)/\partial\varepsilon \\ = {} & \rho^2 - \text{Tr}(\hat{\mathbf{R}}_{\text{SCM}} + 2\varepsilon\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\text{SCM}} + \varepsilon^2\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\text{SCM}}\mathbf{X}_\varepsilon^{-1}) \\ = {} & \rho^2 - \sum_{i=1}^{MN}\left(\kappa_i + \frac{2\varepsilon\kappa_i}{\kappa_{X,i} - \varepsilon} + \frac{\varepsilon^2\kappa_i}{(\kappa_{X,i} - \varepsilon)^2}\right) \\ = {} & \rho^2 - \sum_{i=1}^{MN}\frac{\kappa_{X,i}}{\varepsilon^2} \\ = {} & 0\end{aligned} \tag{39}$$

and the optimal solution of ε satisfies

$$\rho^2 = \sum_{i=1}^{MN}\frac{\kappa_{X,i}}{\varepsilon^2}. \tag{40}$$

At this point, leveraging the property of the rotation-invariant estimator, the closed-form solution of inverse covariance $\mathbf{X}$ is obtained, as summarized in Proposition 2.

PROPOSITION 2 Given $\hat{\mathbf{R}}_{\text{SCM}}$ with its eigenvectors $\mathbf{U}$ and eigenvalues $\{\kappa_i\}_{i=1}^{MN}$, $\{\kappa_i\}_{i=1}^{MN}$ are sorted in ascending order. For any preset uncertainty set size $\rho > 0$, the DRO-based inverse covariance matrix ML estimator for $\mathbf{R}^{-1}$ is constructed via

$$\hat{\mathbf{R}}_{\text{DRO}}^{-1} = \mathbf{U}\text{diag}(\kappa_{X,1}^\star, \kappa_{X,2}^\star, \ldots, \kappa_{X,MN}^\star)\mathbf{U}^H \tag{41}$$

where

$$\kappa_{X,i}^\star = \varepsilon^\star\left(1 + \frac{1}{2}\mathcal{F}_{\varepsilon^\star,\kappa}(\kappa_i)\right),\ i = 1, 2, \ldots, MN \tag{42a}$$

$$\mathcal{F}_{\varepsilon^\star,\kappa}(\kappa_i) = \varepsilon^\star \kappa_i - \sqrt{(\varepsilon^\star \kappa_i)^2 + 4\varepsilon^\star \kappa_i} \tag{42b}$$

$\varepsilon^\star > 0$ denotes the unique positive solution that satisfies equality (40), which can be re-expressed as

$$\left(\rho^2 - \frac{1}{2}\sum_{i=1}^{MN}\kappa_i\right)\varepsilon^\star + \frac{1}{2}\sum_{i=1}^{MN}\sqrt{(\varepsilon^\star \kappa_i)^2 + 4\varepsilon^\star \kappa_i} = MN \tag{43}$$

and $\varepsilon^\star$ can be effectively attained by the bisection method [65]. The solution for $\varepsilon^\star$ and the proof of the existence and uniqueness of $\kappa^\star_{X,i}$ can be seen in Appendix B. ■

As we can see, the closed-form solution of the DRO-based ML estimator functions as a nonlinear shrinkage estimator, as $\hat{\mathbf{R}}^{-1}_{\text{DRO}}$ and $\hat{\mathbf{R}}_{\text{SCM}}$ share common eigenvectors, and $\kappa^\star_{X,i}$ is the nonlinear mapping result of κ_i. More importantly, owing to

$$\frac{\partial \kappa^\star_X}{\partial \kappa} \propto \frac{\partial \mathcal{F}_{\varepsilon^\star,\kappa}(\kappa)}{\partial \kappa} = \varepsilon^\star\left(1 - \frac{\varepsilon^\star \kappa_i + 2}{\sqrt{(\varepsilon^\star \kappa_i)^2 + 4\varepsilon^\star \kappa_i}}\right) < 0 \tag{44}$$

$\mathcal{F}_{\varepsilon^*,\kappa}(\kappa)$ monotonically decreases when $\kappa_i \le \kappa_{i+1}, i = 1, 2, \ldots, MN$ (that is, $\{\kappa_i\}_{i=1}^{MN}$ are sorted in ascending order). Thereby, for any preset $\rho > 0$, $\{(\kappa^\star_{X,i})^{-1}\}_{i=1}^{MN}$ maintain ascending order, that is

$$(\kappa^\star_{X,1})^{-1} \le (\kappa^\star_{X,2})^{-1} \le \cdots \le (\kappa^\star_{X,MN})^{-1}. \tag{45}$$

In other words, $\hat{\mathbf{R}}_{\text{DRO}}$ shrinks the eigenvalues of $\hat{\mathbf{R}}_{\text{SCM}}$ in a way that guarantees order preservation without any extra step, such as pool adjacent violators (PAV) algorithm in [66]. As discussed in Section IV, $\hat{\mathbf{R}}_{\text{DRO}}$ is asymptotic consistent, and

$$\left|\mathcal{F}_{\text{NS}}(\mathbf{s}_t, \hat{\mathbf{R}}^{-1}_{\text{DRO}}, \mathbf{R}) - 1\right| \xrightarrow{\text{p}} 0 \tag{46}$$

as $MN, L \to \infty$ together, where

$$\mathcal{F}_{\text{NS}}(\mathbf{s}_t, \hat{\mathbf{R}}^{-1}, \mathbf{R}) = \frac{(\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}\mathbf{s}_t)^2}{(\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}_t)(\mathbf{s}_t^H \mathbf{R}^{-1}\mathbf{s}_t)} \tag{47}$$

denotes the normalized SCNR. Similar proof of (46) can refer to [41, Appendix D].

REMARK 3 One can observe that the estimation process of $\hat{\mathbf{R}}^{-1}_{\text{DRO}}$ mainly consists of two parts: the eigenvalue decomposition of $\hat{\mathbf{R}}_{\text{SCM}}$ and solving $\varepsilon^\star$ via the bisection method. The corresponding computational complexities are $\mathcal{O}(M^3N^3)$ and $\mathcal{O}(\log_2(1/\delta_c))$, respectively, where δ_c is the convergence parameter. The overall computational complexity of the closed-form solution framework is about $\mathcal{O}(M^3N^3 + \log_2(1/\delta_c))$, which is less than that of employing CVX solvers.

It should be noted that the inverse CNCM estimation accuracy of the proposed DRO-STAP method depends on the statistical distance metric, empirical distribution $\tilde{\mathbb{T}}$, and uncertainty set size ρ. For the given statistical distance metric and empirical distribution (where this article adopts the Wasserstein distance and $\mathcal{CN}(\mathbf{0}, \hat{\mathbf{R}}_{\text{SCM}})$, respectively), a smaller ρ may result in $\Pr\{\mathbb{T} \subseteq \mathbb{D}\} \neq 1$, whereas a larger ρ may lead to an overly conservative estimation, both of which will affect the estimation accuracy. As a consequence, choosing an appropriate uncertainty set size ρ based on training data is attractive for its applications. In the following section, we will briefly introduce a leave-one-out cross-validation (LO2CV) method for selecting uncertainty set size.

E. Choice of the Uncertainty Set Size ρ

Based on the ideal $\mathbf{R}$, the optimal uncertainty set size $\rho^\star$ can be attained by minimizing output clutter power, that is

$$\rho^\star(\mathbf{s}_t) = \arg\min_{\rho>0} \frac{\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)\mathbf{R}\hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)\mathbf{s}_t}{(\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)\mathbf{s}_t)^2}. \tag{48}$$

Since it depends on the ideal $\mathbf{R}$, $\rho^\star$ is unavailable. Given training data $\{\mathbf{z}_l\}_{l=1}^L$, we employ the LO2CV strategy to modify (48), i.e.,

$$\hat{\rho}(\mathbf{s}_t) = \arg\min_{\rho>0} \frac{1}{L}\sum_{l=1}^{L}\left|\frac{\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}_{\text{DRO},l}(\rho)}{\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}_{\text{DRO},l}(\rho)\mathbf{s}_t}\mathbf{z}_l\right|^2 \tag{49}$$

where $\hat{\mathbf{R}}^{-1}_{\text{DRO},l}(\rho)$ denotes the DRO-based inverse covariance matrix ML estimation by using $\{\mathbf{z}_{l'}\}_{l'=1,l'\neq l}^L$. Leveraging the matrix inversion lemma, for the lth split, we can use the following relationship to calculate $\hat{\mathbf{R}}^{-1}_{\text{DRO},l}(\rho)$ with lower computational complexity, that is:

$$\hat{\mathbf{R}}^{-1}_{\text{DRO},l}(\rho) = \hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho) - \frac{\hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)\mathbf{z}_l\mathbf{z}_l^H\hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)}{\mathbf{z}_l^H\hat{\mathbf{R}}^{-1}_{\text{DRO}}(\rho)\mathbf{z}_l - L} \tag{50}$$

and $\hat{\rho}$ can be solved via the numerical search.

For the sake of clarity, we summarize the proposed DRO-STAP method, as shown in Algorithm 1.

F. Computational Complexity

In this section, we briefly conclude the computational complexity of the DRO-STAP method and compare it with several classical STAP methods: fast ML (FML) estimator [22], modified Ledoit–Wolf (LWD2) estimator [41], and random matrix theory (RMT)-based estimator [43]. For DRO-STAP, the computational complexity of the closed form (see Proposition 2) reduces that of the SDP-form (29) from $\mathcal{O}(M^6N^6)$ to $\mathcal{O}(M^3N^3 + \log_2(1/\delta_c))$. For three comparison methods, the main computation lies in the CNCM eigendecomposition, which is $\mathcal{O}(M^3N^3)$. Among these methods, LWD2 yields an extra computational complexity of implementing the PAV algorithm, which is $\mathcal{O}(MN)$.

IV. NUMERICAL EXPERIMENTS

In this section, we verify the performance advantages of the proposed DRO-STAP method through simulated and measured clutter data, respectively. It is important to note that DRO-STAP is derived from the DRO-based ML estimation model. Furthermore, it composes a nonlinear shrinkage

Algorithm 1: DRO-STAP.

Input: Training samples $\{\mathbf{z}_l\}_{l=1}^{L}$

1: Compute the SCM $\hat{\mathbf{R}}_{\text{SCM}}$ via (6) and perform eigen decomposition on $\hat{\mathbf{R}}_{\text{SCM}}$ via (30)
2: Initialize the uncertainty set size ρ
3: Obtain $\varepsilon^{\star}(\rho)$ via (43) and the bisection method [65]
4: Compute $\kappa_{X,i}^{\star}(\rho)$, $i = 1, 2, \ldots, MN$ via (42)
5: Construct $\hat{\mathbf{R}}_{\text{DRO}}^{-1}(\rho)$ via (41)
6: Obtain a suboptimal uncertainty set size $\hat{\rho}$ and corresponding inverse CNCM estimation $\hat{\mathbf{R}}_{\text{DRO}}^{-1}(\hat{\rho})$ via the LO2CV strategy (49) and (50)

Output: Inverse CNCM estimation $\hat{\mathbf{R}}_{\text{DRO}}^{-1}$

estimator, as detailed in Proposition 2. Accordingly, we take the FML estimator [22] (one of the classic regularized ML estimators) and two state-of-the-art nonlinear shrinkage estimators LWD2 [41] and RMT [43] for comparison. For clarity, the working mechanisms of these three estimators are briefly summarized as follows.

1) FML: The FML belongs to a class of regularized ML estimators. It aims to threshold the eigenvalues in SCM from below by the noise power δ_n^2. Generally, δ_n^2 is assumed to be known or estimated.

2) LWD2: The LWD2 thresholds the individualized shrinkage eigenvalues from below by the noise power δ_n^2. Under the large-dimensional assumption, that is, L and MN go to infinity together at the same rate, the LWD2 is claimed to be optimal. Likewise, δ_n^2 is assumed to be known or estimated.

3) RMT: Drawing on the spiked covariance model, the RMT designs a shrinkage framework for clutter-related eigenvalues and estimates noise-related eigenvalues by the noise power δ_n^2. Similarly, the clutter rank c_r and δ_n^2 are assumed to be known or estimated.

One of the effective methods can be used to estimate the clutter rank [27], and δ_n^2 is then assessed via

$$\hat{\delta}_n^2 = \frac{1}{MN - c_r} \sum_{i=1}^{MN-c_r} \kappa_i. \tag{51}$$

Besides, the Oracle estimator [as shown in (32)] and SCM are employed to provide the performance upper and lower bound of these estimators for reference.

A. Simulated Data

In this section, we evaluate the clutter suppression performance and robustness of DRO-STAP by simulation analysis. The simulated clutter data are generated by the signal model as illustrated in Section II-A, and detailed simulation parameters are listed in Table I.

We employ the output SCNR loss and average SCNR loss to quantify the clutter cancellation performance with different estimators, where the SCNR loss is defined as

$$\text{SCNR loss} = \frac{\delta_n^2 \left|\hat{\mathbf{w}}^H \mathbf{s}_t\right|^2}{MN \cdot \hat{\mathbf{w}}^H \mathbf{R} \hat{\mathbf{w}}} \tag{52}$$

where $\hat{\mathbf{w}} = \hat{\mathbf{R}}^{-1}\mathbf{s}_t / \mathbf{s}_t^H \hat{\mathbf{R}}^{-1} \mathbf{s}_t$ denotes the estimated STAP filter and the average SCNR loss is the mean of SCNR loss within entire normalized Doppler frequency range. The following simulation results are attained by averaging 1000 independent Monte Carlo trials.

To begin with, we assess the computational complexity of DRO-STAP. All simulations are conducted on the same software (MATLAB R2021b) and hardware configuration: CPU: Intel Core i7-10700 at 2.90 GHz; RAM: 32 GB. Table II presents the estimated execution time (measured in milliseconds) of SDP-form and closed-form DRO-STAP methods under varying MN, where the SDP-form DRO-STAP is solved by the general-purpose interior point solver SDPT3. As given in Table II, the execution time of SDP-form DRO-STAP significantly exceeds that of closed-form DRO-STAP, particularly as MN increases. The results validate the aforementioned analysis. Unless otherwise stated, we evaluate the performance of closed-form DRO-STAP in the subsequent experiments. Table III compares the runtime of DRO, FML, LWD2, and RMT when $MN = 16$. As we can see from Table III, the runtime of DRO is comparable to that of the other three state-of-the-art methods, with FML exhibiting the shortest execution time. The results indicate

TABLE I
Simulated Airborne Radar System Parameters

Parameters	Values
Platform height	8 km
Platform velocity	150 m/s
Azimuth angle of main lobe	0°
Elevation angle of main lobe	85°
Pulse repetition frequency	2 KHz
Bandwidth	5 MHz
Wavelength	0.3 m
Pulse number	4
Element number	4
Noise power	0 dB
Clutter patch number	361
Clutter-to-noise ratio	40 dB

TABLE II
Execution Time (Measured in Milliseconds) Comparison of Different DRO-STAP Methods Under Varying CNCM Dimension MN

DRO-STAP method	$MN = 4$	$MN = 16$	$MN = 36$
SDP form	1033.245	3636.188	27895.615
Closed form	0.021	0.039	0.171

TABLE III
Execution Time (Measured in Milliseconds) Comparison of Different Methods With $MN = 16$

Method	DRO	FML	LWD2	RMT
Runtime	0.041	0.029	0.051	0.048

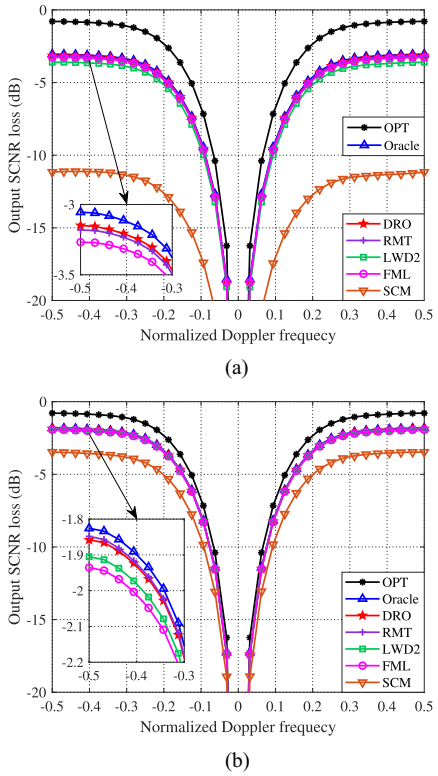

Fig. 1. Output SCNR loss versus normalized Doppler frequency in the ideal case. (a) $L = MN$. (b) $L = 2MN$.

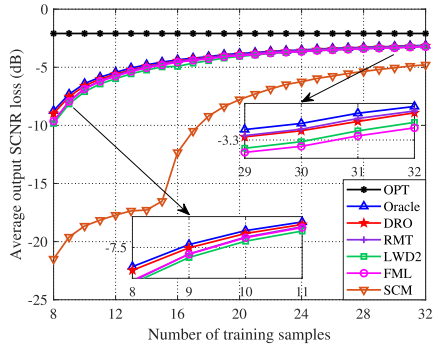

Fig. 2. Average output SCNR loss versus the number of training samples in the ideal case.

that the proposed DRO-STAP is applicable to practical systems.

Drawing on the exact clutter rank, noise power, and uncertainty set size $\rho^{\star}(\mathbf{s}_t)$ [as indicated in (48)], we first verify the performance upper bound of each algorithm. Fig. 1(a) and (b) presents the output SCNR loss curves in the ideal case with $L = MN$ and $L = 2MN$, respectively. The black curve shows the optimal performance with known $\mathbf{R}$. As we can see, in the low sample case ($L = MN$), DRO owns better clutter suppression performance compared with RMT, FML, and LWD2. When the samples go to sufficient ($L = 2MN$), except for SCM, the performance of other estimators is close to Oracle, and RMT performs best at this time.

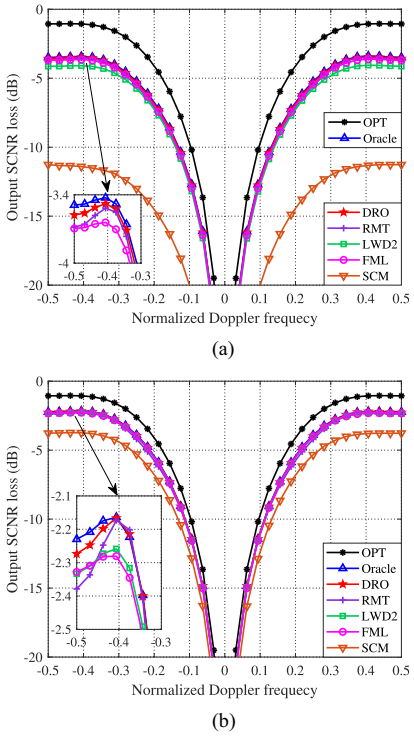

Fig. 3. Output SCNR loss versus normalized Doppler frequency in the nonideal case. (a) $L = MN$. (b) $L = 2MN$.

Fig. 2 depicts the average SCNR loss versus the number of training samples in the ideal case. Four estimators show faster convergence performance than SCM, and DRO owns obvious advantages under small sample support. When the number of samples increases, the performance of DRO is second only to RMT. Meanwhile, the performance of LWD2 gradually becomes better than that of FML. However, LWD2 has performance loss in finite-dimensional cases, thus it is slightly worse than DRO and RMT.

Taking into account the gain-phase error and intrinsic clutter motion (ICM) encountered in practical airborne radar scenarios, Fig. 3(a) and (b) provides the output SCNR loss curves in the nonideal case with $L = MN$ and $L = 2MN$, respectively. Similarly, DRO performs best in the low sample case. In contrast to Fig. 1(b), when samples are sufficient, DRO is better than RMT in the nonideal case. This outcome arises from ICM causing an augmentation in clutter DoFs, thereby impacting the spiked covariance model assumption of RMT and leading to performance degradation.

Fig. 4 shows the average SCNR loss versus the number of training samples in the nonideal case. It is evident that nonideal factors cause the deterioration of the output SCNR by each estimator. As we expected, DRO presents the most robust performance under nonideal conditions. Based on the aforementioned analysis, different from Fig. 2, the clutter suppression performance of RMT is no longer prominent as the number of samples increases.

To analyze the clutter DoFs on the performance of each estimator, Fig. 5 presents the average output SCNR loss versus the carrier velocity v_p under different sample sizes.

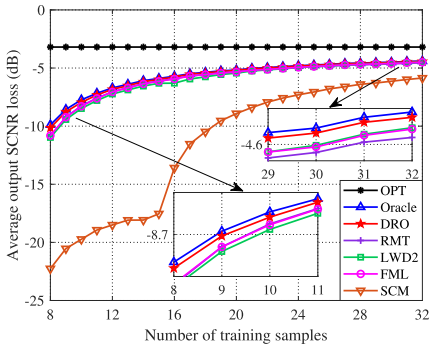

Fig. 4. Average output SCNR loss versus the number of training samples in the nonideal case.

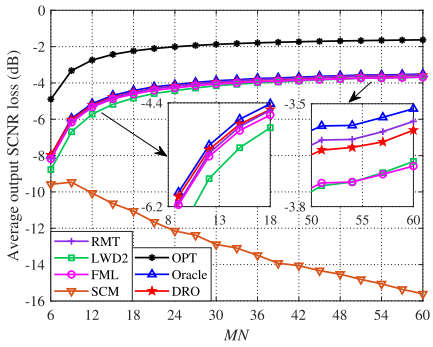

Fig. 6. Average output SCNR loss versus MN with $L = 1.125MN$.

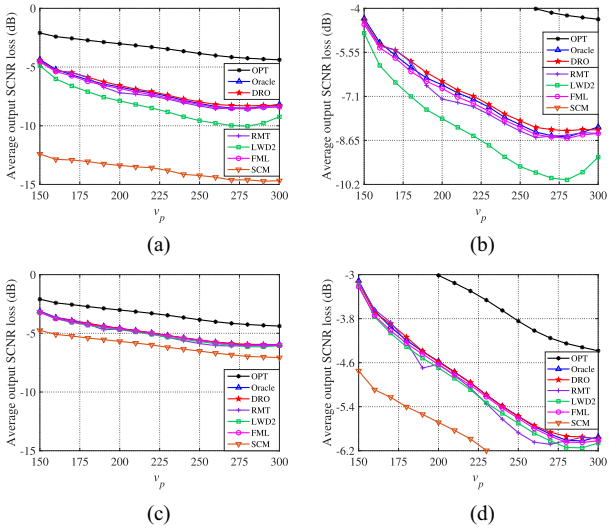

Fig. 5. Average output SCNR loss versus the carrier velocity v_p. (a) $L = MN$. (b) $L = MN$ (zoom in on some areas). (c) $L = 2MN$. (d) $L = 2MN$ (zoom in on some areas).

From Fig. 5, we can see that as v_p increases, the clutter DoFs increase and the clutter suppression performance decreases. Regardless of whether the samples are sufficient or not, the performance advantage and robustness of DRO are satisfactory. Interestingly, DRO and RMT perform better than Oracle in some cases, which is reasonable since Oracle is derived under the asymptotical case, and in finite-dimensional cases, the Frobenius loss metric is not directly relevant to STAP applications [41]. In Fig. 5(d), we can see that DRO is converging toward Oracle when L increases, indicating DRO maybe Oracle consistent. Besides, the performance of FML is stable with the change of v_p, while the performance of RMT fluctuates greatly. The performance of LWD2 is poor at $L = MN$, and the performance improves and is relatively stable at $L = 2MN$.

To further analyze the spiked covariance model and large-dimensional assumption on the performance of each estimator, Fig. 6 plots the average output SCNR loss versus MN. Considering the performance advantage of DRO with low sample support. Meanwhile, from Figs. 2, 4, and 5, it is worth noting that the performance of LWD2 suddenly deteriorates when $L = MN$ [41]. For a fair comparison, these estimators are conducted with $L = 1.125MN$. In Fig. 6, as MN increases, the performance of SCM seriously deteriorates, while the other four estimators are asymptotic consistent with the OPT. In finite-dimensional cases, DRO achieves a better performance than the other estimators. In high-dimensional cases, RMT achieves the best performance due to the additional exploit of the information on clutter DoFs, followed by DRO, and LWD2 gradually surpasses FML. This is reasonable since the spiked covariance model assumption strengthens with $c_r/MN \to 0$ as $MN \to +\infty$. However, as previously analyzed, the performance advantage of RMT and LWD2 heavily depends on this assumption and may not necessarily be robust for uncertain clutter environments. Certainly, this motivates us to undertake further efforts in the future to enhance the performance of DRO in large-dimensional cases.

For practical STAP applications, we take nonideal factors into account and employ the estimated clutter rank, noise power [as shown in (51)], and uncertainty set size $\hat{\rho}(\mathbf{s}_t)$ [as shown in (49)] to construct each estimator with $L = 1.125MN$ samples. Herein, the estimate of noise power is related to clutter rank c_r. Fig. 7 provides the performance analysis on practical STAP application of various estimators. In Fig. 7(a), we can see that the proposed DRO-STAP method still achieves satisfactory clutter suppression performance and performs better than the benchmark estimators. Importantly, in Fig. 7(b), the output performance of RMT, LWD2, and FML changes dramatically with clutter rank c_r. However, for DRO-STAP, the change of uncertainty set size ρ has a relatively little effect on the output performance. The above simulation results validate the effectiveness and robustness of the proposed DRO-STAP method.

To evaluate the performance of the proposed DRO-STAP method under sample mismatch conditions, we analyze the impact of the non-Gaussian setting on the choice of ρ and clutter suppression performance. For a fixed scale parameter v_K, it is well known that the K-distribution gradually reduces to a Gaussian distribution as the shape parameter $\mu_K \to +\infty$ [67]. Therefore, we employ the K-distribution as an example to demonstrate the degree of sample mismatch. Fig. 8(a) illustrates the impact of the choice of ρ on the performance of the proposed method under different

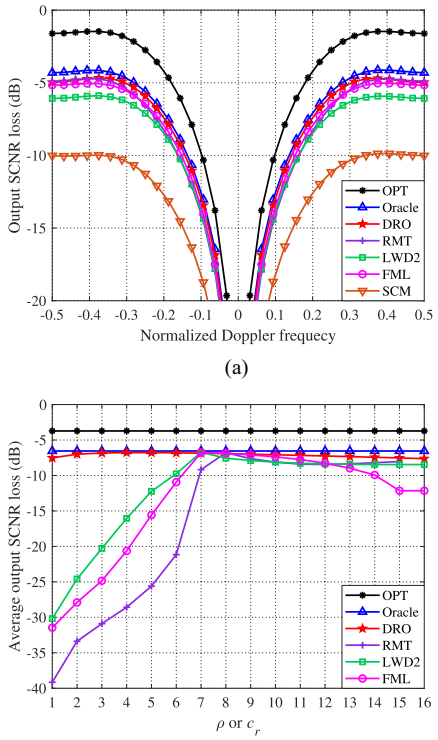

Fig. 7. Performance analysis on practical STAP application of various estimators. (a) Output SCNR loss versus normalized Doppler frequency with estimated parameters. (b) Average output SCNR loss versus ρ or c_r.

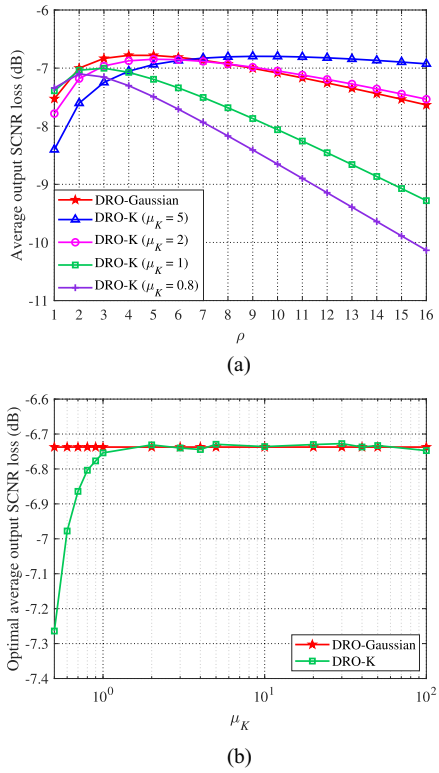

Fig. 8. Performance analysis on DRO-STAP for mismatched training samples. (a) Average output SCNR loss versus ρ. (b) Optimal average output SCNR loss versus μ_K.

TABLE IV
Measured Airborne Radar System Parameters

Parameters	Values
Platform height	3988 m
Platform velocity	107 m/s
Azimuth angle of main lobe	0°
Elevation angle of main lobe	89.4°
Pulse repetition frequency	250 Hz
Pulsewidth	0.0002 s
Wavelength	0.67 m
Pulse number	16
Element number	16
Number of range cells	7590

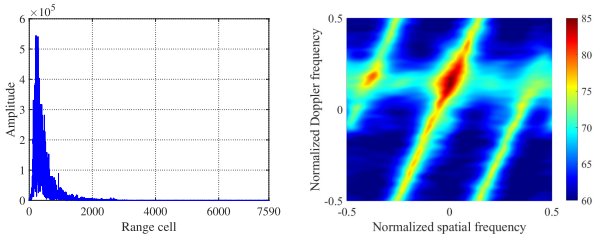

Fig. 9. Measured airborne radar clutter data. (a) Signal amplitude versus the range cell at pulse 1 of element 1 after MF. (b) Capon spectrum.

degrees of sample mismatch (i.e., $\mu_K = 5, 2, 1, 0.8$, labeled DRO-K) when $v_K = 1$. It can be observed that as the degree of sample mismatch increases (i.e., as μ_K decreases), ρ^* gradually shifts leftward. Meanwhile, the influence of the choice of ρ on the output performance becomes significant, leading to a slight reduction in the robustness of DRO-STAP. Compared to the sample matched scenario (labeled DRO-Gaussian), the optimal clutter suppression performance of DRO-STAP decreases under sample mismatch conditions. To further analyze the impact of sample mismatch on the clutter suppression performance of DRO-STAP, Fig. 8(b) shows the curves of optimal average output SCNR loss versus μ_K, where $\mu_K \in [0.5, 100]$. It is evident that as μ_K increases, the degree of sample mismatch gradually decreases, and the output performance progressively approaches that of the sample matched scenario.

B. Measured Data

In this section, we confirm the target detection performance and robustness of DRO-STAP by measured clutter data. The clutter data are collected by a side-looking airborne pulsed-Doppler radar, and detailed system parameters are listed in Table IV.

Fig. 9(a) and (b) presents the signal amplitude versus the range cell at pulse 1 of element 1 after MF and the Capon spectrum of the measured clutter data, respectively. In Fig. 9, we can see that the clutter power distribution versus the range cell is quite nonuniform, and the clutter spectrum is folded due to the low pulse repetition frequency, resulting in increased clutter DoFs.

For convenience, we select subaperture data consisting of $M = 6$ consecutive pulses and $N = 4$ consecutive

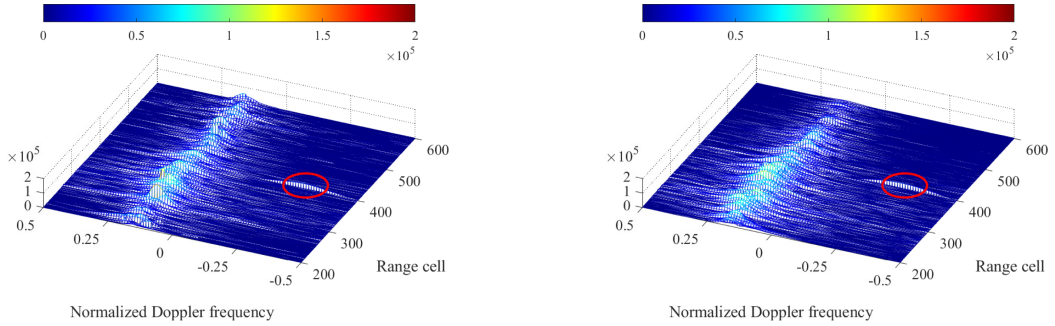

Fig. 10. STAP output range–Doppler spectrum (amplitude) of the proposed DRO-STAP method. (a) $L = MN$. (b) $L = 2MN$.

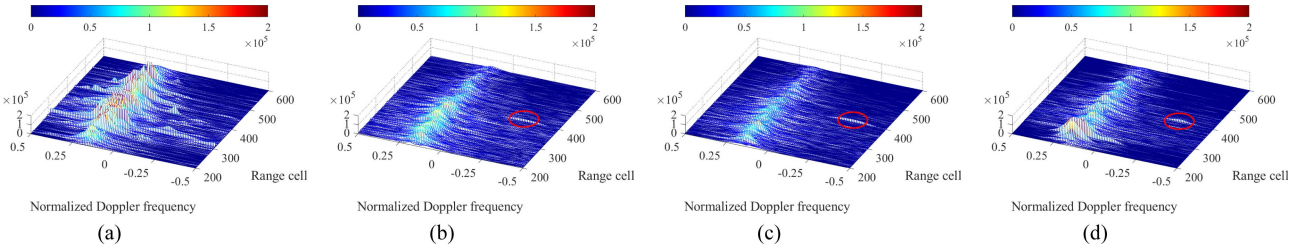

Fig. 11. STAP output range–Doppler spectrum (amplitude) of various methods with $L = MN$. (a) SCM. (b) RMT. (c) FML. (d) LWD2.

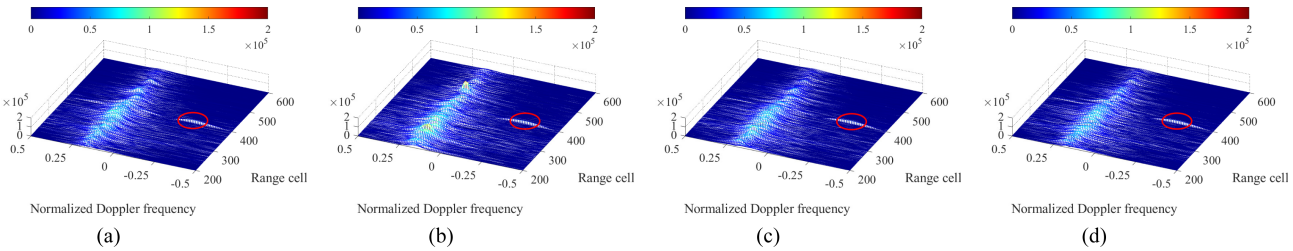

Fig. 12. STAP output range–Doppler spectrum (amplitude) of various methods with $L = 2MN$. (a) SCM. (b) RMT. (c) FML. (d) LWD2.

elements for subsequent target detection. To improve the reliability of statistical analysis, the processing window is from the 200th range cell to the 600th range cell, including 401 range cells. A synthetic target with normalized spatial frequency 0 and normalized Doppler frequency -0.3 is added in the 400th range cell. For each CUT, four samples symmetrically adjacent to it are regarded as the guard cells. Meantime, we utilize the covariance similarity metric [68] to select L near-homogeneous samples from $2.5MN$ candidate range cells.

Employing the sliding window method, Figs. 10–12 present the STAP output range–Doppler spectrum (amplitude) of various methods with $L = MN, 2MN$, respectively. From Figs. 10(a) and 11, when $L = MN$, SCM cannot effectively detect the target due to insufficient training samples. The proposed DRO-STAP method achieves better target detection performance, followed by FML. RMT exhibits a higher false alarm, and LWD2 shows relatively poor clutter suppression performance in some range–Doppler regions. From Figs. 10(b) and 12, when $L = 2MN$, all methods can effectively detect the target and reduce false alarms, and LWD2 exhibits improved clutter suppression performance in the specific range–Doppler regions, as shown in Fig. 11(d). In both cases, the proposed DRO-STAP method presents satisfied target detection performance, demonstrating its effectiveness and robustness in practical application.

Considering that the adaptive matched filter (AMF) [1] owns the constant false alarm rate property, we further use AMF for a quantitative comparison of various methods similar to [69], having

$$P_{\mathrm{AMF}} = \frac{\left|\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}\mathbf{z}\right|^2}{\mathbf{s}_t^H \hat{\mathbf{R}}^{-1}\mathbf{s}_t}. \tag{53}$$

Figs. 13–15 present the normalized AMF output power versus the range cell of the proposed method and other estimators with $L = MN, 2MN$, respectively. If the target can be detected (which has the largest output power within the processing window, as indicated by the circle), we assess the target detection performance by two metrics: the peak sidelobe level (PSL), defined as the output power difference between the target-detected range cell and the range cell

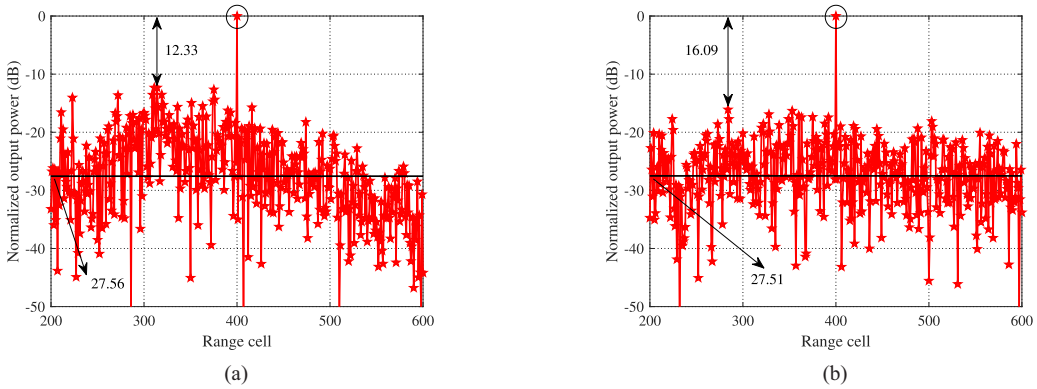

Fig. 13. Normalized AMF output power versus the range cell of the proposed DRO-STAP method. (a) $L = MN$. (b) $L = 2MN$.

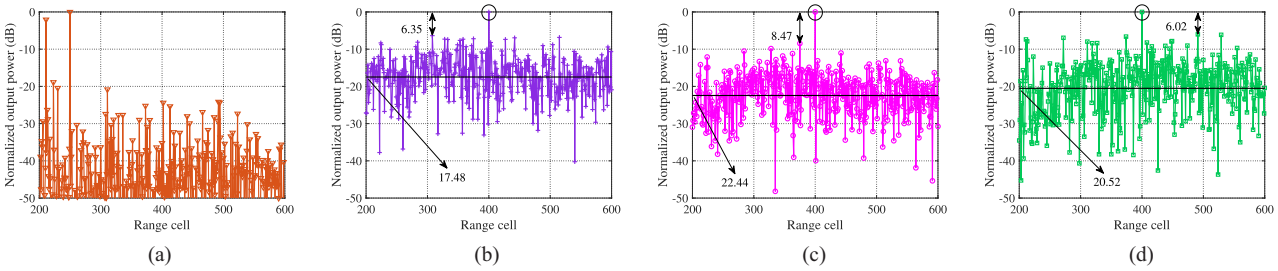

Fig. 14. Normalized AMF output power versus the range cell of various methods with $L = MN$. (a) SCM. (b) RMT. (c) FML. (d) LWD2.

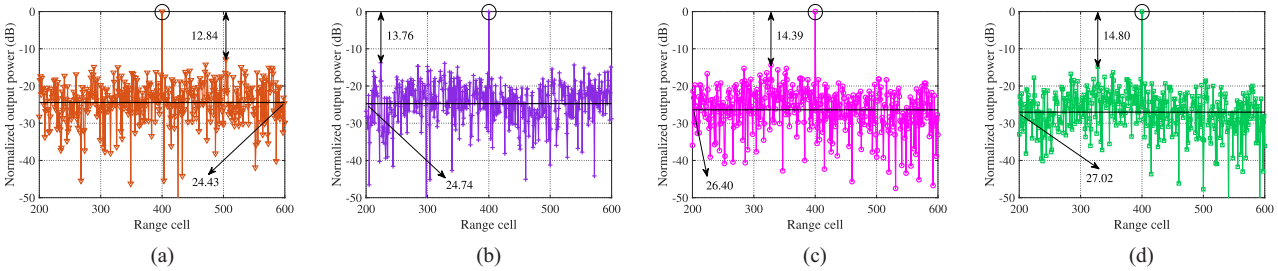

Fig. 15. Normalized AMF output power versus the range cell of various methods with $L = 2MN$. (a) SCM. (b) RMT. (c) FML. (d) LWD2.

with the second largest output power except the guard range cell, and the statistical output SCNR, defined as the ratio of the target output power to the average output power of all range cells except both the target-detected and guard range cells.

From Figs. 13(a) and 14, SCM cannot detect the target since the training samples are insufficient. RMT exhibits comparable PSL performance to LWD2; however, its output SCNR is approximately 3 dB lower than that of LWD2. Although LWD2 has a lower average clutter output power, the PSL is high, resulting in a large false alarm. FML outperforms both RMT and LWD2 in terms of PSL and output SCNR. In contrast, the proposed DRO method achieves the best overall performance, with PSL and output SCNR approximately 4 and 5 dB better than those of FML, respectively. From Figs. 13(b) and 15, SCM can detect the target when the samples become sufficient, but its PSL and output SCNR performance are both worse than those of the other four methods. At this point, the PSL performance of the proposed method improves by 3.76 dB, while the enhancement in output SCNR is not significant. The PSL and output SCNR of the other three methods all improve, with the LWD2 exhibiting the most significant enhancement. However, its PSL and output SCNR are still 1.29 and 0.49 dB worse than those of the proposed method, respectively. The above experimental results confirm the effectiveness and robustness of the proposed DRO-STAP method.

V. CONCLUSION

In this article, we present a DRO-STAP algorithm to estimate the inverse CNCM for Gaussian clutter with low sample support. In this algorithm, the robust estimator is derived from the DRO-based ML estimation problem with the Wasserstein metric. We reformulate the resulting

nonconvex problem into the SDP problem and provide a closed-form solution for low computational complexity. The existence and uniqueness of the solution have been proved theoretically. For practical DRO-STAP applications, a suboptimal choice of the uncertainty set size is discussed by leveraging the LO2CV method. The resulting closed-form estimator utilizes the rotation-equivariant property and composes a nonlinear shrinkage estimator. It shrinks the sample eigenvalues in a way that guarantees order preservation without any extra step. Compared with other state-of-the-art estimators, abundant numerical experiments verify the effectiveness and robustness of DRO-STAP, especially for small sample, nonideal, and finite-dimensional cases.

For future work, we consider extending DRO-STAP to non-Gaussian clutter models and try to enhance its performance in large-dimensional cases while maintaining robustness.

APPENDIX A
PROOF OF PROPOSITION 1

To begin with, a useful lemma will be introduced for the subsequent proof.

LEMMA 1 (TRACE INEQUALITIES) For $\mathbf{A}, \mathbf{B} \in \mathbb{H}^{MN}$, $\boldsymbol{\kappa}_A$ and $\boldsymbol{\kappa}_B$, respectively, denote their eigenvalue vectors. Then, the following inequality holds:

$$(\boldsymbol{\kappa}_A^{\downarrow})^T \boldsymbol{\kappa}_B^{\uparrow} \le \mathrm{Tr}(\mathbf{AB}) \le (\boldsymbol{\kappa}_A^{\downarrow})^T \boldsymbol{\kappa}_B^{\downarrow} \tag{54}$$

and the equality in (54) holds when $\mathbf{A}$ and $\mathbf{B}$ commute.

Recall that the inner maximization subproblem (25) can be rewritten as

$$\max_{\mathbf{Y}\succeq\mathbf{0}} \mathrm{Tr}\left(\mathbf{Y}\mathbf{X}_\varepsilon + 2\varepsilon\left(\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\mathbf{Y}\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right). \tag{55}$$

In the following, our proof based on the assumption $\hat{\mathbf{R}}_{\mathrm{SCM}} \succ \mathbf{0}$, even though $\hat{\mathbf{R}}_{\mathrm{SCM}}$ is singular, the same conclusion still holds since the problem (20) is continuous in $\hat{\mathbf{R}}_{\mathrm{SCM}}$. For more details, one can refer to [56].

For briefness, let us define $\mathbf{Y}_R = (\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}}\mathbf{Y}\hat{\mathbf{R}}_{\mathrm{SCM}}^{\frac{1}{2}})^{\frac{1}{2}}$ and $\mathbf{X}_R = \hat{\mathbf{R}}_{\mathrm{SCM}}^{-\frac{1}{2}}\mathbf{X}_\varepsilon\hat{\mathbf{R}}_{\mathrm{SCM}}^{-\frac{1}{2}}$, then the problem (55) can be simplified to

$$\max_{\mathbf{Y}_R\succeq\mathbf{0}} \mathrm{Tr}(\mathbf{Y}_R^2\mathbf{X}_R) + 2\varepsilon\mathrm{Tr}(\mathbf{Y}_R). \tag{56}$$

Reformulating $\mathbf{Y}_R$ and $\mathbf{X}_R$ regarding their eigenvalue decomposition is $\mathbf{Y}_R = \mathbf{U}_{\tilde{Y}}\mathrm{diag}(\boldsymbol{\kappa}_{\tilde{Y}})\mathbf{U}_{\tilde{Y}}^H$ and $\mathbf{X}_R = \mathbf{U}_{\tilde{X}}\mathrm{diag}(\boldsymbol{\kappa}_{\tilde{X}})\mathbf{U}_{\tilde{X}}^H$, where $\mathbf{U}_{\tilde{Y}}$ and $\mathbf{U}_{\tilde{X}}$, respectively, are the eigenvector matrices of $\mathbf{Y}_R$ and $\mathbf{X}_R$, $\boldsymbol{\kappa}_{\tilde{Y}}$ and $\boldsymbol{\kappa}_{\tilde{X}}$ are the corresponding eigenvalue vectors. Both $\boldsymbol{\kappa}_{\tilde{Y}}$ and $\boldsymbol{\kappa}_{\tilde{X}}$ are arranged in descending order.

Using the fact of trace inequalities in Lemma 1, we find

$$\begin{aligned}
&\max_{\mathbf{Y}_R\succeq\mathbf{0}} \mathrm{Tr}(\mathbf{Y}_R^2\mathbf{X}_R) + 2\varepsilon\mathrm{Tr}(\mathbf{Y}_R)\\
&= \max_{\mathbf{Y}_R\succeq\mathbf{0}} \mathrm{Tr}(\mathbf{U}_{\tilde{Y}}\mathrm{diag}(\boldsymbol{\kappa}_{\tilde{Y}}^2)\mathbf{U}_{\tilde{Y}}^H\mathbf{U}_{\tilde{X}}\mathrm{diag}(\boldsymbol{\kappa}_{\tilde{X}})\mathbf{U}_{\tilde{X}}^H + 2\varepsilon\mathbf{Y}_R)\\
&\le \max_{\boldsymbol{\kappa}_{\tilde{Y}}\succeq\mathbf{0}} (\boldsymbol{\kappa}_{\tilde{Y}}^2)^T\boldsymbol{\kappa}_{\tilde{X}} + 2\varepsilon\mathbf{1}_{MN}^T\boldsymbol{\kappa}_{\tilde{Y}}
\end{aligned} \tag{57}$$

and the equality in (57) holds when $\mathbf{U}_{\tilde{X}} = \mathbf{U}_{\tilde{Y}}$.

It is clear that (57) is bounded only when $\boldsymbol{\kappa}_{\tilde{X}} \prec \mathbf{0}$ (that is, $\mathbf{X}_\varepsilon \prec \mathbf{0}$); otherwise, it is unbounded. Leveraging this constraint, the optimal solution of $\mathbf{Y}_R$ can be computed via the first-order optimality condition

$$\frac{\partial\mathrm{Tr}(\mathbf{Y}_R^2\mathbf{X}_R + 2\varepsilon\mathbf{Y}_R)}{\partial\mathbf{Y}_R} = \mathbf{X}_R^*\mathbf{Y}_R^* + \mathbf{Y}_R^*\mathbf{X}_R^* + 2\varepsilon\mathbf{I}_{MN} = 0. \tag{58}$$

Thus, the unique maximizer $\mathbf{Y}_R^\star$ and the maximum of (56), respectively, are $\mathbf{Y}_R^\star = -\varepsilon\mathbf{X}_R^{-1}$ and $-\varepsilon^2\mathrm{Tr}(\mathbf{X}_R^{-1}) = -\varepsilon^2\mathrm{Tr}(\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}})$. Correspondingly, we can get $\mathbf{Y}^\star = \varepsilon^2\mathbf{X}_\varepsilon^{-1}\hat{\mathbf{R}}_{\mathrm{SCM}}\mathbf{X}_\varepsilon^{-1}$.

This completes the proof.

APPENDIX B
PROOF OF PROPOSITION 2

Recall that the first-order optimality condition of $\mathbf{X}$ satisfies (38), that is

$$\kappa_{X,i}^2 - (2\varepsilon + \varepsilon^2\kappa_i)\kappa_{X,i} + \varepsilon^2 = 0 \quad \forall i \le MN. \tag{59}$$

If there are no constraints on $\kappa_{X,i}$, it is easy to attain two candidate solutions of $\kappa_{X,i}$ by the quadratic equation formula, i.e.,

$$\kappa_{X,i}^{\star_1} = \varepsilon^\star\left(1 + \frac{1}{2}\left(\varepsilon^\star\kappa_i - \sqrt{(\varepsilon^\star\kappa_i)^2 + 4\varepsilon^\star\kappa_i}\right)\right) \tag{60a}$$

$$\kappa_{X,i}^{\star_2} = \varepsilon^\star\left(1 + \frac{1}{2}\left(\varepsilon^\star\kappa_i + \sqrt{(\varepsilon^\star\kappa_i)^2 + 4\varepsilon^\star\kappa_i}\right)\right) \tag{60b}$$

where $\varepsilon^\star > 0$ satisfies equality (40). That is to say, $\varepsilon^\star$ is a positive solution of the following algebraic equation:

$$\left(\rho^2 - \frac{1}{2}\sum_{i=1}^{MN}\kappa_i\right)\varepsilon + \frac{1}{2}\sum_{i=1}^{MN}\sqrt{(\varepsilon\kappa_i)^2 + 4\varepsilon\kappa_i} - MN = 0. \tag{61}$$

Let us define $\mathcal{F}_{\rho,\kappa,\varepsilon}(\varepsilon)$ as the left-hand side of (61). For any $\rho > 0$, since $\mathcal{F}_{\rho,\kappa,\varepsilon}(0) < 0$, $\mathcal{F}_{\rho,\kappa,\varepsilon}(+\infty) > 0$, and $\partial\mathcal{F}_{\rho,\kappa,\varepsilon}(\varepsilon)/\partial\varepsilon > 0$, $\mathcal{F}_{\rho,\kappa,\varepsilon}(\varepsilon)$ is monotonically increasing in $\varepsilon > 0$, and $\varepsilon^\star$ is the unique positive solution of (61). Meanwhile, we can calculate the upper and lower bounds of $\varepsilon^\star$ similar to [56], giving

$$\begin{aligned}
&\frac{MN\left(MN\kappa_{\max} + 2\rho^2 - \sqrt{M^2N^2\kappa_{\max}^2 + 4MN\rho^2\kappa_{\max}}\right)}{2\rho^4}\\
&\le \varepsilon^\star \le \min\left\{\frac{MN}{\rho^2}, \frac{1}{\rho}\sqrt{\sum_{i=1}^{MN}\frac{1}{\kappa_i}}\right\}.
\end{aligned} \tag{62}$$

Thereby, $\varepsilon^\star$ can be effectively obtained by the bisection method [65].

As for choosing $\kappa_{X,i}^{\star_1}$ or $\kappa_{X,i}^{\star_2}$, since $\varepsilon^\star > 0$ and

$$\varepsilon^\star - \kappa_{X,i}^{\star_1} = \varepsilon^\star\left(\frac{1}{2}\left(\sqrt{(\varepsilon^\star\kappa_i)^2 + 4\varepsilon^\star\kappa_i} - \varepsilon^\star\kappa_i\right)\right) > 0 \tag{63}$$

$\kappa_{X,i}^{\star_1}$ obeys the constraint $\mathbf{X}_\varepsilon \prec \mathbf{0}$ in (28) while $\kappa_{X,i}^{\star_2}$ does not; thus, $\kappa_{X,i}^{\star_1}$ is the unique solution of (59) for now. The issue then evolves to whether $\kappa_{X,i}^{\star_1}$ satisfies $\mathbf{X} \succ \mathbf{0}$ in (28).

Let us define $x_i = \varepsilon^\star \kappa_i$ and $\mathcal{F}_x(x) = x - \sqrt{x^2 + 4x}$, then $\kappa_{X,i}^{\star_1}$ can be rewritten as

$$\kappa_{X,i}^{\star_1} = \varepsilon^\star \left(1 + \frac{1}{2}\mathcal{F}_x(x_i)\right). \tag{64}$$

If $\kappa_i = 0$, then $\kappa_{X,i}^{\star} = \varepsilon^\star > 0$.

If $\kappa_i > 0$, then $x_i > 0$, and the derivative of $\mathcal{F}_x(x)$ satisfies

$$\frac{\partial \mathcal{F}_x(x)}{\partial x} = 1 - \frac{x+2}{\sqrt{x^2+4x}} < 0 \tag{65}$$

Thus, $\mathcal{F}_x(x)$ is monotonically decreasing in $x > 0$. Let $y = \frac{1}{x}$, we find

$$\begin{aligned}
&\lim_{x\to+\infty}\left(1 + \frac{1}{2}\mathcal{F}_x(x)\right) = 1 + \lim_{x\to+\infty}\frac{1}{2}x\left(1 - \sqrt{1+\frac{4}{x}}\right)\\
&\triangleq 1 + \lim_{x\to+\infty}\frac{1}{2}x - \lim_{y\to0^+}\frac{1}{2}x\sqrt{1+4y}\\
&\approx 1 + \lim_{x\to+\infty}\frac{1}{2}x\left(2y^2 - 2y\right)\\
&= 1 + \lim_{x\to+\infty}\left(\frac{1}{x} - 1\right)\\
&\to 0^+
\end{aligned} \tag{66}$$

where the approximate equality in (66) holds by leveraging the second-order Taylor expansion of $\sqrt{1+4y}$ at $y = 0$, i.e.,

$$\sqrt{1+4y} \approx 1 + 2y - 2y^2. \tag{67}$$

Then, it is natural to attain

$$\lim_{x\to+\infty}\kappa_{X,i}^{\star_1} \propto \lim_{x\to+\infty}\left(1 + \frac{1}{2}\mathcal{F}_x(x)\right) > 0 \tag{68}$$

and

$$\kappa_{X,i}^{\star_1} > 0 \quad \forall i \le MN. \tag{69}$$

According to (63) and (69), we can see that $\kappa_{X,i}^{\star_1}$ satisfies all constraints in (28), thereby $\kappa_{X,i}^{\star_1}$ is the unique reasonable solution of (38).

This completes the proof.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] W. L. Melvin, "A STAP overview," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 19–35, Jan. 2004.

[2] P. Huang, Z. Zou, X.-G. Xia, X. Liu, and G. Liao, "A novel dimension-reduced space–time adaptive processing algorithm for spaceborne multichannel surveillance radar systems based on spatial–temporal 2-D sliding window," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5109721.

[3] Z. Esmaeilbeig, K. V. Mishra, and M. Soltanalian, "Space-time adaptive processing for radars in connected and automated vehicular platoons," in *Proc. ICASSP 2024-2024 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2024, pp. 13056–13060.

[4] Z. Liu, S. Zhu, J. Xu, L. Lan, X. He, and G. Liao, "Enhanced parameters estimation with STAP-Based PA-FDA dual mode radar," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 60, no. 5, pp. 6925–6940, Oct. 2024.

[5] Y. Li, Y. Wang, Y. Zhang, B. Liu, and H. Chen, "A method for calculating the optimal velocity search step size for airborne three-channel SAR adaptive clutter suppression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5712–5720, 2024.

[6] Y. Li, J. Chen, and J. Zhu, "A new ground accelerating target imaging method for airborne CSSAR," *IEEE Geosci. Remote. Sens. Lett.*, vol. 21, 2024, Art. no. 4013305.

[7] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electr. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[8] W. L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.

[9] R. C. DiPietro, "Extended factored space-time processing for airborne radar systems," in *Proc. 26th Asilomar Conf. Signals, Syst. Comput.*, 1992, pp. 425–430.

[10] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 660–670, Jul. 1994.

[11] Y.-L. Wang, J.-W. Chen, Z. Bao, and Y.-N. Peng, "Robust space-time adaptive processing for airborne radar in nonhomogeneous clutter environments," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 39, no. 1, pp. 70–81, Jan. 2003.

[12] W. Zhang, Z. He, J. Li, H. Liu, and Y. Sun, "A method for finding best channels in beam-space post-Doppler reduced-dimension STAP," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 50, no. 1, pp. 254–264, Jan. 2014.

[13] N. Cui, K. Duan, K. Xing, and Z. Yu, "Beam-space reduced-dimension 3D-STAP for nonside-looking airborne radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 3506505.

[14] K. Duan, H. Xu, H. Yuan, H. Xie, and Y. Wang, "Reduced-DOF three-dimensional STAP via subarray synthesis for nonsidelooking planar array airborne radar," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 56, no. 4, pp. 3311–3325, Aug. 2020.

[15] A. Haimovich and M. Berin, "Eigenanalysis-based space-time adaptive radar: Performance analysis," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 33, no. 4, pp. 1170–1179, Oct. 1997.

[16] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.

[17] X. Wang, E. Aboutanios, and M. G. Amin, "Reduced-rank STAP for slow-moving target detection by antenna-pulse selection," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1156–1160, Aug. 2015.

[18] D. Song, S. Chen, Q. Feng, F. Xi, and Z. Liu, "Subspace-based STAP in airborne FDA radar for homogeneous clutter under range ambiguity," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 60, no. 3, pp. 2909–2921, Jun. 2024.

[19] T. K. Sarkar et al., "A deterministic least-squares approach to space-time adaptive processing (STAP)," *IEEE Trans. Antennas Propag.*, vol. 49, no. 1, pp. 91–103, Jan. 2001.

[20] D. Cristallini and W. Burger, "A robust direct data domain approach for STAP," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1283–1294, Mar. 2012.

[21] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[22] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[23] K. Gerlach and M. L. Picciolo, "Airborne/spacebased radar STAP using a structured covariance matrix," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 39, no. 1, pp. 269–281, Jan. 2003.

[24] A. L. Kraay and A. B. Baggeroer, "A physically constrained maximum-likelihood method for snapshot-deficient adaptive array processing," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4048–4063, Aug. 2007.

[25] G. Ginolhac, P. Forster, F. Pascal, and J. P. Ovarlez, "Exploiting persymmetry for low-rank space time adaptive processing," *Signal Process.*, vol. 97, pp. 242–251, 2014.

[26] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 50, no. 1, pp. 501–515, Jan. 2014.

[27] B. Kang, V. Monga, M. Rangaswamy, and Y. Abramovich, "Expected likelihood approach for determining constraints in covariance estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2139–2156, Oct. 2016.

[28] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, Jul. 2016.

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

[30] X. Zhu, J. Li, and P. Stoica, "Knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 47, no. 2, pp. 1325–1336, Apr. 2011.

[31] M. Riedl and L. C. Potter, "Multimodel shrinkage for knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 54, no. 5, pp. 2601–2610, Oct. 2018.

[32] Y. Xiong, W. Xie, H. Li, and X. Gao, "Colored-loading factor optimization for airborne KA-STAP radar," *IEEE Sensors J.*, vol. 23, no. 19, pp. 23317–23326, Oct. 2023.

[33] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.

[34] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5640–5651, Nov. 2014.

[35] E. Ollila and D. E. Tyler, "Regularized *m*-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.

[36] A. De Maio, L. Pallotta, J. Li, and P. Stoica, "Loading factor estimation under affine constraints on the covariance eigenvalues with application to radar target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1269–1283, Jun. 2019.

[37] E. Ollila and E. Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2707–2719, May 2019.

[38] D. Song, Q. Feng, S. Chen, F. Xi, and Z. Liu, "Space–time adaptive processing using deep neural network-based shrinkage algorithm under small training samples," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 9697–9703, Dec. 2023.

[39] O. Ledoit and M. Wolf, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *Ann. Statist.*, vol. 40, no. 2, pp. 1024–1060, 2012.

[40] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *Ann. Statist.*, vol. 48, no. 5, pp. 3043–3065, 2020.

[41] B. D. Robinson, R. Malinas, and A. O. Hero, "Space-time adaptive detection at low sample support," *IEEE Trans. Signal Process.*, vol. 69, pp. 2939–2954, 2021.

[42] S. Jain, V. Krishnamurthy, M. Rangaswamy, B. Kang, and S. Gogineni, "Radar clutter covariance estimation: A nonlinear spectral shrinkage approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 7640–7653, Dec. 2023.

[43] D. Song, S. Chen, H. Li, and F. Xi, "Space-time adaptive processing via random matrix theory for finite training samples," *IEEE Sensors J.*, vol. 23, no. 7, pp. 7334–7344, 2023.

[44] Z. Yang, R. C. de Lamare, and X. Li, "L_1-regularized STAP algorithms with a generalized sidelobe canceler architecture for airborne radar," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 674–686, Feb. 2012.

[45] K. Duan, Z. Wang, W. Xie, H. Chen, and Y. Wang, "Sparsity-based STAP algorithm with multiple measurement vectors via sparse Bayesian learning strategy for airborne radar," *IET Signal Process.*, vol. 11, no. 5, pp. 544–553, 2017.

[46] Z. Wang, W. Xie, K. Duan, and Y. Wang, "Clutter suppression algorithm based on fast converging sparse Bayesian learning for airborne radar," *Signal Process.*, vol. 130, pp. 159–168, 2017.

[47] D. Wang, T. Wang, W. Cui, and X. Zhang, "A clutter suppression algorithm via enhanced sparse Bayesian learning for airborne radar," *IEEE Sensors J.*, vol. 23, no. 10, pp. 10900–10911, May 2023.

[48] Y. Wang, J. Wang, X. Zhang, J. Li, and Z. He, "Knowledge-aided multi-dictionary block sparsity-aware STAP for airborne polarimetric conformal array radar," *Signal Process.*, vol. 224, 2024, Art. no. 109585.

[49] N. Cui, K. Xing, Z. Yu, and K. Duan, "Tensor-based sparse recovery space-time adaptive processing for large size data clutter suppression in airborne radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 907–922, Apr. 2023.

[50] K. Duan, H. Chen, W. Xie, and Y. Wang, "Deep learning for high-resolution estimation of clutter angle-Doppler spectrum in STAP," *IET Radar, Sonar Navigat.*, vol. 16, no. 2, pp. 193–207, 2022.

[51] Z. Hangui, F. Weike, F. Cunqian, Z. Bo, and L. Fuyu, "Deep unfolding based space-time adaptive processing method for airborne radar," *J. Radars*, vol. 11, no. 4, pp. 676–691, 2022.

[52] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A review," 2019, *arXiv:1908.05659*.

[53] D. Kuhn, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, "Wasserstein distributionally robust optimization: Theory and applications in machine learning," in *Proc. INFORMS Tut. Operations Res.*, 2019, pp. 130–166.

[54] I. Yang, "Wasserstein distributionally robust stochastic control: A data-driven approach," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3863–3870, Aug. 2021.

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

[56] V. A. Nguyen, D. Kuhn, and P. Mohajerin Esfahani, "Distributionally robust inverse covariance estimation: The Wasserstein shrinkage estimator," *Operations Res.*, vol. 70, no. 1, pp. 490–515, 2022.

[57] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.

[58] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, 2002.

[59] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, pp. 234–251, 1987.

[60] A. Hakobyan and I. Yang, "Wasserstein distributionally robust control of partially observable linear stochastic systems," *IEEE Trans. Automat. Contr.*, 2024.

[61] M. Gelbrich, "On a formula for the L2 Wasserstein metric between measures on Euclidean and Hilbert spaces," *Mathematische Nachrichten*, vol. 147, no. 1, pp. 185–203, 1990.

[62] A. Shapiro, "On duality theory of conic linear problems," *Nonconvex Optim. Appl.*, vol. 57, pp. 135–165, 2001.

[63] M. Deng, Z. Cheng, L. Wu, B. Shankar, and Z. He, "One-bit ADCs/DACs based MIMO radar: Performance analysis and joint design," *IEEE Trans. Signal Process.*, vol. 70, pp. 2609–2624, 2022.

[64] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.

[65] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.

[66] O. Ledoit and M. Wolf, "Working paper 264: Direct nonlinear shrinkage estimation of large-dimensional covariance matrices," Working Paper, Tech. Rep. 264, 2017.

[67] K. Ward, R. Tough, and S. Watts, *Sea Clutter: Scattering, the K Distribution and Radar Performance*. 2nd ed., London, U.K.: IET, 2013.
[68] Y. Wu, T. Wang, J. Wu, and J. Duan, "Training sample selection for space-time adaptive processing in heterogeneous environments," *IEEE Geosci. Remote. Sens. Lett.*, vol. 12, no. 4, pp. 691–695, Apr. 2015.
[69] S. Han, C. Fan, and X. Huang, "A novel STAP based on spectrum-aided reduced-dimension clutter sparse recovery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 213–217, Feb. 2017.

Yalong Wang (Graduate Student Member, IEEE) was born in Anhui, China, in 2000. He received the B.S. degree in communication engineering from the School of Information and Communication Engineering, North University of China, Taiyuan, China, in 2021. He is currently working toward the Ph.D. degree in information and communication engineering with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include array signal processing, space–time adaptive processing, and radar adaptive target detection.

Xuejing Zhang (Member, IEEE) was born in Hebei, China. He received the B.S. degree in electrical engineering from Huaqiao University, Xiamen, China, in 2011, the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2014, and the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019.

From 2017 to 2019, he was a Visiting Student with the University of Delaware, Newark, DE, USA. He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include array signal processing and integer programming, with applications to radar and communications.

Dr. Zhang was the recipient of the Young Scientists Award for Excellence in Scientific Research by the International Union of Radio Science in 2020 and the Prize for Excellent Doctor Degree Dissertation of the Chinese Institute of Electronics in 2020.

Zhihang Wang (Member, IEEE) was born in Zhangping, China, in 1993. He received the B.S. degree in electronic information engineering and Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016 and 2022, respectively.

He is currently an Assistant Research Fellow and a Postdoctoral Fellow with the School of Information and Communication Engineering, UESTC. His research interests include radar target detection, array signal processing, compressive sensing, and machine learning.

Jun Li was born in Sichuan, China, in 1977. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2000, 2003, and 2009, respectively.

He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include multi-input multi-output radar, cognitive radar, array signal processing, and space–time adaptive processing.

Zishu He (Senior Member, IEEE) was born in Sichuan, China, in 1962. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), in 1984, 1988, and 2000, respectively.

He is currently a Professor in signal and information processing with the School of Information and Communication Engineering, UESTC. His research interests include array signal processing, space–time adaptive processing, digital beamforming, the theory of multiple-input multiple-output (MIMO) communication and MIMO radar, and interference cancellation.