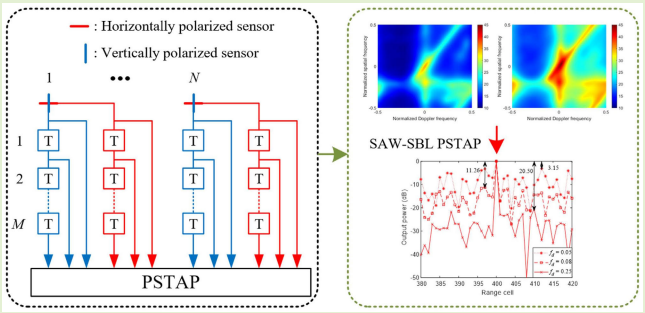
Fast Converging and Controllable Structure-Aware Clutter Suppression Method for Airborne Polarimetric Array Radar

Yalong Wang, *Graduate Student Member, IEEE*, Jiaheng Wang, Xuejing Zhang, *Member, IEEE*, Jun Li, and Zishu He, *Senior Member, IEEE*

***Abstract*—Polarimetric space-time adaptive processing (PSTAP) significantly enhances the ability to detect low-speed targets for airborne early warning radar. However, incorporating diverse polarization sensor data poses challenges: it expands the snapshot dimension and increases clutter heterogeneity. Therefore, the performance of PSTAP may suffer due to the inaccurate estimation of the clutter-plus-noise covariance matrix (CNCM) with finite samples. Leveraging the Kronecker product structure of clutter, statistical framework-based Kronecker estimators can reduce sample requirements while maintaining the clutter suppression performance. But for low-speed target detection, we theoretically illustrate the limitations of this type of estimator. While sparse recovery (SR) space-time adaptive processing (STAP) methods can achieve satisfactory CNCM estimation with very few samples, they cannot be directly applied to PSTAP. In this article, we propose a structure-aware (SAW) sparse Bayesian learning (SBL) algorithm for PSTAP, named SAW-SBL PSTAP. By exploiting the independence between the polarization domain and the space-time domain, along with the intrinsic sparsity of clutter in the angle-Doppler plane, we model a block SR problem and develop a fast and controllable learning framework. This framework alternately updates the noise power, polarization covariance matrix, and clutter space-time power, resulting in precise CNCM estimation. Both simulated and measured data experiments verify the effectiveness and robustness of the proposed method, particularly in enhancing detection performance for low-speed targets.**



***Index Terms*— Covariance matrix estimation, heterogeneous clutter, polarimetric array, polarimetric space-time adaptive processing (PSTAP), sparse recovery (SR).**

I. Introduction

DETECTING weak targets submerged in the strong clutter background is the main task of moving platform array radar systems [1], [2]. Space-time adaptive processing (STAP) technology jointly employs the spatial and temporal domains that can effectively suppress the clutter and enhance targets [3], [4], [5]. But for low-speed or tangentially moving targets, it is difficult to discriminate the target from clutter only in the angle-Doppler domain, resulting in a significant deterioration in target detection performance. Unlike single-polarized radar, polarimetric array radar equips multiple polarization sensors and thus owns the capability to perceive polarimetric information about the target and the clutter [6]. Significant research indicates that leveraging polarization diversity can capture the scattering information of the scene of interest and enhance the target detection performance, particularly in scenarios where the discrimination between targets and clutter in the angle-Doppler domain is minimal [7], [8], [9], [10], [11], [12].

For the research on polarimetric STAP (PSTAP), Showman et al. [13] introduced the concept of PSTAP and examined the performance of two PSTAP architectures. However, its primary objective is to overcome the small apertures of STAP radar. Wu et al. [14] theoretically analyzed the advantages of PSTAP compared to STAP in detecting

Received 20 December 2024; revised 15 January 2025 and 22 January 2025; accepted 2 February 2025. Date of publication 11 February 2025; date of current version 14 March 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62101101, Grant 62031007, and Grant 62231006; in part by Sichuan Science and Technology Program under Grant 2024NSFSC1433; and in part by the Peng Cheng Shang Xue Education Fund under Grant XY2021602. The associate editor coordinating the review of this article and approving it for publication was Prof. Danilo Orlando. *(Corresponding author: Xuejing Zhang.)*

The authors 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; jiaheng_wang_0823@163.com; zhangxuejing@uestc.edu.cn; lijunsc@uestc.edu.cn; zshe@uestc.edu.cn).

Digital Object Identifier 10.1109/JSEN.2025.3538793

slowly moving targets, and show that factors such as clutter degree of polarization, polarization difference between target and clutter, and the input clutter-to-noise ratio (CNR) are the key points affecting the performance of PSTAP. However, this work mainly aims at optimal PSTAP, but we have no access to the ideal clutter-plus-noise covariance matrix (CNCM) in practice. To further enhance the detection performance on slowly moving targets, our research group presents a transmitter polarization optimization method for PSTAP and derives an iterative solution framework by the block majorization-minimization criteria [15]. Similarly, this approach also assumes that CNCM is known a priori.

The analysis above indicates that accurate CNCM estimation is key to fully leveraging the performance of PSTAP. Because the polarimetric radar can operate in different polarization channels, the dimension of received data has been expanded. The well-known Reed, Mallet, and Brennan rule [16] proves that the output signal-to-clutter-plus-noise ratio (SCNR) loss is not more than 3 dB when the number of independent and identically distributed (IID) training data is not less than twice the system degrees of freedom (DoFs). This poses a great challenge for polarimetric radar since external objects often display varying scattering properties [17]. Meanwhile, the clutter environment is usually heterogeneous in actual scenarios [18]. These factors result in limited samples. Thus, improving the clutter suppression performance at low sample support is a key issue in PSTAP applications.

To solve the problem of limited samples, lots of state-of-the-art approaches have been proposed to improve the accuracy of CNCM estimation. Knowledge-aided (KA) STAP [19], [20], [21] leverages available prior knowledge to improve performance, but its effectiveness is limited by the reliability of the prior model and the choice of loading factors. Reduced-dimensional (RD) [22], [23], [24] STAP converts full-dimensional data into low-dimensional data by selecting several channels in the angle-Doppler domain, thereby reducing the need for the number of samples. How to extend these RD methods to polarimetric radar needs further research. Reduced-rank (RR) [25], [26], [27] STAP projects the full-dimensional data into the low-dimensional clutter subspace, but as the clutter degree of polarization increases accompanied by input CNR decreases, or for nonside looking arrays and conformal arrays, accurately determining the clutter rank becomes challenging, thus reducing the practicality of these methods. Leveraging the prior information of data covariance, various structured estimators are developed to reduce the number of unknown parameters and enhance estimation accuracy [28], [29], [30], [31], [32], [33], [34], [35]. Leveraging two sets of training data, i.e., classic training data surrounding the cell under test (CUT) with an additional training set, methods presented in [36], [37], [38], and [39] can also enhance the CNCM estimation accuracy in sample starved scenarios. Meanwhile, a series of shrinkage algorithms [40], [41], [42] are derived to improve estimation accuracy. However, their performance remains constrained in severely heterogeneous clutter environments.

Recently, sparse recovery (SR) technology has gained attention across various fields for its high-precision signal reconstruction capabilities [43]. Numerous SR-based STAP algorithms have been proposed to improve the CNCM estimation performance with few samples. These algorithms are usually classified into three categories: greedy algorithms, convex or nonconvex optimization algorithms, and sparse Bayesian learning (SBL) algorithms. Classic greedy algorithms such as orthogonal matching pursuit [44] are simple to implement and have lower computational complexity, but the effectiveness in STAP is seriously affected by input CNR and the correlation among dictionary atoms. Convex/nonconvex optimization algorithms [45], [46], [47] employ the relaxation model to obtain a satisfactory recovery accuracy, but the setting of regularization parameters is a common issue for these methods. On the contrary, SBL algorithms [48], [49], [50], [51], [52] have gained the most attention due to their excellent recovery accuracy, insensitivity to atomic correlation, and self-adjusting parameters. However, these algorithms are mainly developed based on STAP. How to apply SR technology to PSTAP is of great significance.

In this article, we theoretically prove the performance upper bound of Kronecker estimators and highlight their limitations in low-speed target detection. To address these challenges and enhance detection performance for low-speed targets, we integrate SR technology into PSTAP and propose a structure-aware SBL (SAW-SBL) algorithm with limited samples. We begin by analyzing the sparse representation of PSTAP data and identifying the limitations of existing SR-STAP algorithms when applied to PSTAP, laying the groundwork for the development of SAW-SBL. Then, we exploit the Kronecker structure of the clutter polarization-space-time covariance matrix and the inherent sparsity of clutter in the complete angle-Doppler plane to formulate a block SR problem. To solve this, we derive a fast converging, controllable learning framework that alternately estimates three key parameters: noise power, the clutter polarization covariance matrix, and clutter space-time power. Finally, the accurately estimated CNCM is applied to PSTAP. The effectiveness of SAW-SBL is validated through experiments with both simulated and real PSTAP data, demonstrating superior performance in terms of convergence, clutter suppression, and particularly, low-speed target detection.

The remainder of this article is organized as follows. Section II introduces the signal model, optimal PSTAP principle, and the CNCM estimation problem. Section III discusses the question of SR-based PSTAP and analyzes the details of the proposed SAW-SBL PSTAP method. Section IV provides the experimental results to verify the effectiveness of the proposed method. Section V concludes this article.

Notations: In this article, $\mathbf{A}^{-1}$ and $\mathbf{A}_{i,j}$ are the inverse and the element in the ith row and jth column of matrix $\mathbf{A}$, respectively. diag($\mathbf{a}$) is the diagonalization of vector $\mathbf{a}$. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $\otimes$, and $\mathbb{E}\{\cdot\}$ are the transpose, conjugate transpose, Kronecker product, and statistical expectation operators, respectively. $\mathbf{I}_N$ is the N-dimensional identity matrix. rank$(\cdot)$ and Tr$(\cdot)$ are the rank and trace of a matrix, respectively. $|\cdot|$ is the absolute value of a scalar or the determinant of a matrix. $\|\cdot\|_2$, $\|\cdot\|_{2,0}$, and $\|\cdot\|_{\mathrm{F}}$ are the l_2 norm, $l_{2,0}$ mixed norm, and Frobenius norm,

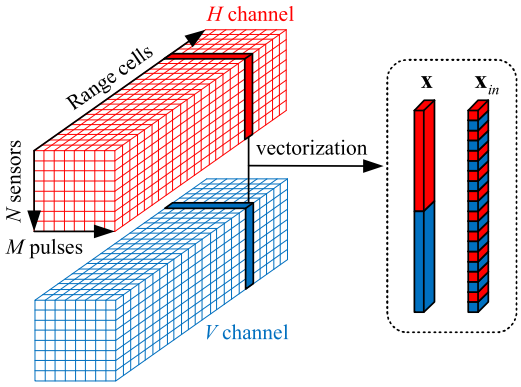


Fig. 1. Structure of PSTAP data.

respectively. $\propto$ means that the quantities on its both sides are directly proportional.

II. Preliminaries

A. Signal Model and Optimal PSTAP

We consider a side-looking pulsed Doppler airborne radar equipped with a uniform linear array (ULA) consisting of N diversely polarized antennas. Each antenna owns two orthogonal sensors to measure both the horizontal and vertical polarization components (E_H and E_V) of the echo signal. M pulses with constant pulse repetition frequency (PRF) f_r are transmitted during a coherent processing interval (CPI). For the sake of illustration, we employ $N_p = 2$ polarization channels similar to [14]. Note that the proposed method is also applicable to $N_p = 3$ or $N_p = 4$. The received digital baseband signal $\mathbf{x} \in \mathbb{C}^{N_p NM}$ of a single range cell can be modeled as follows:

$$\mathbf{x} = \alpha_t \mathbf{a}_{\text{pst}}^t + \mathbf{x}_c + \mathbf{x}_n \triangleq \left[\mathbf{x}_H^{\text{T}}, \mathbf{x}_V^{\text{T}}\right]^{\text{T}} \tag{1}$$

where $\mathbf{x}_H = [x_{H,1}, x_{H,2}, \ldots, x_{H,MN}]^{\text{T}} \in \mathbb{C}^{MN}$ and $\mathbf{x}_V = [x_{V,1}, x_{V,2}, \ldots, x_{V,MN}]^{\text{T}} \in \mathbb{C}^{MN}$ denote the data belonging to the H channel (horizontally polarized sensors) and the V channel (vertically polarized sensors), respectively. α_t and $\mathbf{a}_{\text{pst}}^t$ are the complex reflection coefficient and polarization-space-time steering vector of the potential target, respectively. For notational simplicity, let $N_D = N_p NM$. $\mathbf{x}_c$ denotes the clutter signal. $\mathbf{x}_n$ represents the zero-mean complex Gaussian white noise with variance matrix $\sigma^2 \mathbf{I}_{N_D}$, and σ^2 is the noise power.

Herein, we define $\mathbf{x}$ in (1) as a deinterleaved form. We also define the interleaved form $\mathbf{x}_{\text{in}} \in \mathbb{C}^{N_D}$, having

$$\mathbf{x}_{\text{in}} \triangleq [x_{H,1}, x_{V,1}, x_{H,2}, x_{V,2}, \ldots, x_{H,MN}, x_{V,MN}]^{\text{T}}. \tag{2}$$

For clarity, the relationship between $\mathbf{x}$ and $\mathbf{x}_{\text{in}}$ is illustrated in Fig. 1. Generally, the received clutter signal $\mathbf{x}_c$ can be formulated as the accumulation of all ambiguous ranges and clutter patches in the same range cell [3], i.e.,

$$\mathbf{x}_c = \sum_{i=1}^{N_r}\sum_{j=1}^{N_c} \alpha_{i,j} \mathbf{a}_{\text{pst}}^{i,j} = \sum_{i=1}^{N_r}\sum_{j=1}^{N_c} \alpha_{i,j} \mathbf{p}_{i,j} \otimes \mathbf{a}_{\text{st}}^{i,j} \tag{3}$$

where N_r and N_c are the numbers of range ambiguity and uniformly divided clutter patches, respectively. $\alpha_{i,j}$ is the complex reflection coefficient of the i, jth clutter patch. $\mathbf{a}_{\text{pst}}^{i,j} = \mathbf{p}_{i,j} \otimes \mathbf{a}_{\text{st}}^{i,j}$, $\mathbf{p}_{i,j} \in \mathbb{C}^{N_p}$, and $\mathbf{a}_{\text{st}}^{i,j} \in \mathbb{C}^{NM}$ denote the polarization-space-time, polarization, and space-time steering vectors of the corresponding clutter patch, respectively.

The normalized polarization covariance matrix $\mathbf{R}_p \in \mathbb{C}^{N_p \times N_p}$ is utilized to represent the polarization correlation characteristic of the H channel and the V channel clutter, i.e.,

$$\mathbf{R}_p = \mathbb{E}\{\mathbf{p}_{i,j}\mathbf{p}_{i,j}^{\text{H}}\} = \begin{bmatrix} 1 & \sqrt{\chi_c}|\rho_c|e^{-j\bar{\eta}_c} \\ \sqrt{\chi_c}|\rho_c|e^{j\bar{\eta}_c} & \chi_c \end{bmatrix} \tag{4}$$

where χ_c, ρ_c, and $\bar{\eta}_c$ denote the power ratio, complex correlation coefficient, and statistical average phase difference between the two polarization channels, respectively.

Note that independence between the polarization and the space-time responses [8], the clutter polarization-space-time covariance matrix can be written as follows:

$$\begin{aligned} \mathbf{R}_{\text{pst}} &= \mathbb{E}\{\mathbf{x}_c\mathbf{x}_c^{\text{H}}\} = \mathbf{R}_p \otimes \mathbf{R}_{\text{st}} \\ &= \mathbf{R}_p \otimes \sum_{i=1}^{N_r}\sum_{i=1}^{N_c} \mathbb{E}\left\{|\alpha_{i,j}|^2\right\} \mathbf{a}_{\text{st}}^{i,j}\left(\mathbf{a}_{\text{st}}^{i,j}\right)^{\text{H}} \end{aligned} \tag{5}$$

where $\mathbf{R}_{\text{st}}$ denotes the clutter space-time covariance matrix, then the ideal target-free CNCM can be given as follows:

$$\mathbf{R} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\text{H}}\} = \mathbf{R}_p \otimes \mathbf{R}_{\text{st}} + \sigma^2 \mathbf{I}_{N_D}. \tag{6}$$

The well-known optimal PSTAP weight vector can be computed via the minimum variance distortionless response criterion [5], having

$$\mathbf{w} = \tau \mathbf{R}^{-1} \mathbf{a}_{\text{pst}}^t \tag{7}$$

where $\tau = 1/(\mathbf{a}_{\text{pst}}^t)^{\text{H}} \mathbf{R}^{-1} \mathbf{a}_{\text{pst}}^t$ denotes the normalized factor.

B. Problem Formulation and CNCM Estimation

Note that it is difficult to obtain the optimal PSTAP since we have no access to the ideal CNCM $\mathbf{R}$. For practical application, $\mathbf{R}$ is estimated through L IID training samples $\{\mathbf{x}_l\}_{l=1}^L$, one of the classical statistical methods is the sample covariance matrix (SCM) approach, i.e.,

$$\mathbf{R}_{\text{SCM}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{x}_l \mathbf{x}_l^{\text{H}}. \tag{8}$$

However, homogeneous training samples are limited in actual clutter scenarios. Meantime, the clutter rank or clutter DoFs of PSTAP ($\mathbf{R}_{\text{pst}}$) is higher than or equal to (only when the clutter is completely polarized wave, i.e., $|\rho_c| = 1$) that of STAP ($\mathbf{R}_{\text{st}}$), which is given by the following equation:

$$\begin{aligned} \text{rank}(\mathbf{R}_{\text{st}}) &\le \text{rank}(\mathbf{R}_{\text{pst}}) \\ &= \text{rank}(\mathbf{R}_p \otimes \mathbf{R}_{\text{st}}) = \text{rank}(\mathbf{R}_p)\text{rank}(\mathbf{R}_{\text{st}}) \\ &\le N_p \text{rank}(\mathbf{R}_{\text{st}}). \end{aligned} \tag{9}$$

Therefore, the clutter suppression performance is seriously degraded with limited samples, accurately estimating CNCM at low sample support is the prerequisite for taking advantage of the performance of PSTAP.

Leveraging the Kronecker structure of $\mathbf{R}_{\text{pst}} = \mathbf{R}_p \otimes \mathbf{R}_{\text{st}}$, several statistical framework-based Kronecker estimators [8],

[29], [33] are developed to relieve the sample requirement. Note that these estimators generally consider a noise-free (high CNR) signal model, but for PSTAP applications, the noise component is not negligible. Thus, the CNCM $\mathbf{R}$ in (6) does not strictly maintain the Kronecker structure. The following proposition illustrates the limitation of Kronecker estimators.

Proposition 1: By employing the Kronecker product structure of $\mathbf{R}_{\text{pst}}$. Statistical framework-based Kronecker estimators approximate $\mathbf{R} \approx \mathbf{R}_p \otimes \mathbf{R}_{\text{st}}$ to reduce the sample requirement. Importantly, this relaxation deteriorates the detection performance of PSTAP for low-speed targets.

Proof: See Appendix A. ■

Meanwhile, it is still challenging for statistical framework-based estimators to achieve satisfactory CNCM estimation performance in severely nonuniform and nonstationary clutter environments, especially, the polarization characteristics of the clutter also vary greatly within the range.

III. SAW-SBL PSTAP Algorithm

In this section, to accurately reconstruct CNCM and ensure its matrix structure, we introduce the proposed SR algorithm for PSTAP with very few samples. Specifically, we first study the sparse representation problem of PSTAP data and then examine the issues that arise when directly applying existing SR-STAP methods to PSTAP. Finally, we introduce a SAW SBL method with fast and controllable convergence for PSTAP and discuss its computational complexity.

A. Analysis of SR-Based PSTAP Algorithm

The SR technology reconstructs CNCM by utilizing the inherent sparse characteristics of the clutter. For PSTAP, the L deinterleaved samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{C}^{N_D \times L}$ can be rewritten as the multiple measurement vectors model

$$\mathbf{X} = (\mathbf{I}_{N_p} \otimes \mathbf{D}_{\text{st}})\boldsymbol{\Psi} = \mathbf{D}\boldsymbol{\Psi} \tag{10}$$

where $\mathbf{D}_{\text{st}} \in \mathbb{C}^{MN \times K}$ is the overcomplete STAP dictionary similar to [48], K is the number of grid points. $\mathbf{D} = \mathbf{I}_{N_p} \otimes \mathbf{D}_{\text{st}} \in \mathbb{C}^{N_D \times N_p K}$ denotes the deinterleaved PSTAP dictionary. For notational simplicity, let $N_K = N_p K$, and $\boldsymbol{\Psi} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_L] \in \mathbb{C}^{N_K \times L}$ represents the sparse coefficient matrix corresponding to both H channel and V channel clutter.

In SR algorithms, the estimation of $\boldsymbol{\Psi}$ can be formulated as follows:

$$\hat{\boldsymbol{\Psi}} = \arg\min_{\boldsymbol{\Psi}} \|\boldsymbol{\Psi}\|_{2,0}, \quad \text{s.t. } \|\mathbf{X} - \mathbf{D}\boldsymbol{\Psi}\|_{\text{F}}^2 \le \xi \tag{11}$$

where ξ is a given threshold related to the noise power [53]. After obtaining the estimation of $\boldsymbol{\Psi}$ and σ^2, if we reconstruct the CNCM similar to existing SR-based STAP methods, i.e.,

$$\begin{aligned}\hat{\mathbf{R}}_{\text{SR-D}} &= \mathbf{D}\text{diag}\left(\frac{1}{L}\sum_{l=1}^{L}|\hat{\boldsymbol{\gamma}}_l|^2\right)\mathbf{D}^{\text{H}} + \hat{\sigma}^2\mathbf{I}_{N_D} \\ &= \mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{\text{H}} + \hat{\sigma}^2\mathbf{I}_{N_D}\end{aligned} \tag{12}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix, and we define this method as SR-diagonal (SR-D)-based PSTAP. Note that $\hat{\mathbf{R}}_{\text{SR-D}}$ here has a structural loss compared to the $\mathbf{R}$ in (6).

Specifically, $\mathbf{R}$ in (6) can be rewritten as follows:

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_p \otimes \mathbf{R}_{\text{st}} + \sigma^2\mathbf{I}_{N_D} \\ &= \left(\mathbf{I}_{N_p}\mathbf{R}_p\mathbf{I}_{N_p}\right) \otimes \left(\mathbf{D}_{\text{st}}\boldsymbol{\Lambda}_{\text{st}}\mathbf{D}_{\text{st}}^{\text{H}}\right) + \sigma^2\mathbf{I}_{N_D} \\ &= \left(\mathbf{I}_{N_p} \otimes \mathbf{D}_{\text{st}}\right)\left(\mathbf{R}_p \otimes \boldsymbol{\Lambda}_{\text{st}}\right)\left(\mathbf{I}_{N_p} \otimes \mathbf{D}_{\text{st}}^{\text{H}}\right) + \sigma^2\mathbf{I}_{N_D} \\ &= \mathbf{D}\left(\mathbf{R}_p \otimes \boldsymbol{\Lambda}_{\text{st}}\right)\mathbf{D}^{\text{H}} + \sigma^2\mathbf{I}_{N_D} \\ &= \mathbf{D}\mathbf{B}\mathbf{D}^{\text{H}} + \sigma^2\mathbf{I}_{N_D}\end{aligned} \tag{13}$$

where $\boldsymbol{\Lambda}_{\text{st}}$ is a diagonal matrix, representing the clutter space-time power spectrum [48]. Since the clutter of the H channel and V channel has a certain correlation, i.e., $|\rho_c| \neq 0$, it is obvious that $\boldsymbol{\Lambda}$ and $\mathbf{B} = \mathbf{R}_p \otimes \boldsymbol{\Lambda}_{\text{st}}$ are unequal in matrix structure. The loss of matrix structure is essentially an inaccuracy in estimating the clutter polarization covariance matrix $\mathbf{R}_p$, thus leading to a deterioration in the detection performance of low-speed targets, which is detailedly discussed in Section IV.

Different from (12), [54] estimates $\mathbf{R}_p$ and $\mathbf{R}_{\text{st}}$ separately, and then reconstructs $\mathbf{R}$ based on the Kronecker structure, i.e.,

$$\begin{aligned}\hat{\mathbf{R}}_{\text{SR-K}} &= \eta\hat{\mathbf{R}}_p \otimes \left(\mathbf{D}_{\text{st}}\text{diag}\left(\hat{\mathbf{p}}_{\text{st}}\right)\mathbf{D}_{\text{st}}^{\text{H}}\right) + \hat{\sigma}^2\mathbf{I}_{N_D} \\ &= \eta\hat{\mathbf{R}}_p \otimes \hat{\mathbf{R}}_{\text{st}} + \hat{\sigma}^2\mathbf{I}_{N_D}\end{aligned} \tag{14}$$

where η is the regularization parameter used to balance the clutter power and noise power. Due to the low polarization dimension, the traditional SCM method is employed to estimate $\mathbf{R}_p$. $\hat{\mathbf{p}}_{\text{st}} \in \mathbb{C}^K$ and $\hat{\sigma}^2$ are the estimated clutter space-time power and noise power by employing SR technology. We define this method as SR-Kronecker (SR-K)-based PSTAP.

Remark 1: If $\hat{\mathbf{R}}_p$, $\hat{\mathbf{p}}_{\text{st}}$, and $\hat{\sigma}^2$ are all estimated precisely, we can set $\eta = 1$, but it is difficult for SR-K to achieve an accurate estimation of parameters $\mathbf{R}_p$, $\mathbf{p}_{\text{st}}$, and σ^2 in practical applications. Thus, the output performance can be improved by adjusting η. When η is set too large, the estimated noise power decreases. Leveraging the analysis in Appendix A, the performance of the low-speed target detection deteriorates in this case; When η is set too small, the estimated clutter power decreases, and the overall clutter suppression performance deteriorates.

Meanwhile, the preset η is not suitable for time-varying clutter environments. It is necessary to study the robust CNCM estimation method for PSTAP.

B. CNCM Estimation Using SAW-SBL

Note that adaptively estimating parameters $\mathbf{R}_p$, $\boldsymbol{\Lambda}_{\text{st}}$, and σ^2 at few sample support to reconstruct CNCM by (6) is the key to ensuring the performance of PSTAP. To achieve this, we first need to arrange L samples $\{\mathbf{x}_l\}_{l=1}^L$ into the interleaved form $\{\mathbf{x}_{\text{in},l}\}_{l=1}^L$, the same as (2). Drawing on the interleaved signal model, we can easily derive a diagonal block SR-based solution framework to update parameters $\mathbf{R}_p$, $\boldsymbol{\Lambda}_{\text{st}}$, and σ^2 alternately. $\mathbf{x}_{\text{in},l}, l = 1, 2, \ldots, L$ can be rewritten as follows:

$$\mathbf{x}_{\text{in},l} = (\mathbf{D}_{\text{st}} \otimes \mathbf{I}_{N_p})\boldsymbol{\beta}_l = \mathbf{D}_{\text{in}}\boldsymbol{\beta}_l \tag{15}$$

where $\mathbf{D}_{\text{in}} = \mathbf{D}_{\text{st}} \otimes \mathbf{I}_{N_p} \in \mathbb{C}^{N_D \times N_K}$ is the interleaved PSTAP dictionary, and $\boldsymbol{\beta}_l, l = 1, 2, \ldots, L$ is the sparse coefficient

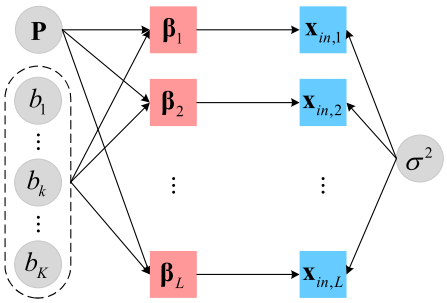


Fig. 2. Hierarchical SAW-SBL model graph.

vector. In this case, the CNCM can be reexpressed as follows:

$$\begin{aligned}\mathbf{R}_{\text{in}} &= \mathbb{E}\{\mathbf{x}_{\text{in},l}\mathbf{x}_{\text{in},l}^{\text{H}}\} = \mathbf{R}_{\text{st}} \otimes \mathbf{R}_p + \sigma^2\mathbf{I}_{N_D} \\ &= \left(\mathbf{D}_{\text{st}}\boldsymbol{\Lambda}_{\text{st}}\mathbf{D}_{\text{st}}^{\text{H}}\right) \otimes \left(\mathbf{I}_{N_p}\mathbf{R}_p\mathbf{I}_{N_p}\right) + \sigma^2\mathbf{I}_{N_D} \\ &= \mathbf{D}_{\text{in}} \underbrace{\left(\boldsymbol{\Lambda}_{\text{st}} \otimes \mathbf{R}_p\right)}_{\mathbf{C}} \mathbf{D}_{\text{in}}^{\text{H}} + \sigma^2\mathbf{I}_{N_D} \\ &= \mathbf{D}_{\text{in}}\mathbb{E}\{\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}\}\mathbf{D}_{\text{in}}^{\text{H}} + \sigma^2\mathbf{I}_{N_D} \end{aligned} \tag{16}$$

where $\mathbf{C} = \mathbb{E}\{\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}\} = \boldsymbol{\Lambda}_{\text{st}} \otimes \mathbf{R}_p$ is a block diagonal matrix. Based on interleaved form samples, we can estimate $\mathbf{R}_p$, $\boldsymbol{\Lambda}_{\text{st}}$, and σ^2 alternately via the proposed SAW-SBL method. It is important to note that even with the interleaved signal model, SR-D PSTAP in (12) does not estimate $\mathbf{R}_p$ or does not estimate $\mathbf{R}_p$ accurately, the structural loss problem remains and has not been alleviated.

To maintain the matrix structure of $\mathbf{R}_{\text{in}}$, we assume that $\boldsymbol{\beta}_l$'s follow zero-mean Gaussian sparse prior distribution with covariance $\boldsymbol{\Sigma}$, having:

$$p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}),\ l = 1, 2, \ldots, L \tag{17}$$

$$\boldsymbol{\Sigma} = \text{diag}(\mathbf{b}) \otimes \mathbf{P} = \text{diag}(b_1, \ldots, b_k, \ldots, b_K) \otimes \mathbf{P} \tag{18}$$

where $\mathbf{b} = [b_1, \ldots, b_k, \ldots, b_K]^{\text{T}}$ is the hyperparameter vector that controls the sparsity of clutter space-time profiles. The positive semidefinite matrix $\mathbf{P} \in \mathbb{C}^{N_p \times N_p}$ indicates the correlation property of H channel and V channel clutter. Furthermore, σ^2 also follows the Gaussian distribution assumption. In a nutshell, the hierarchical SAW-SBL model graph of the proposed method is depicted in Fig. 2.

Leveraging the Bayesian rule, we know that the posterior distribution $p(\boldsymbol{\beta}_l|\mathbf{x}_{\text{in},l}; \mathbf{b}, \mathbf{P}, \sigma^2), l = 1, 2, \ldots, L$ also obeys the complex Gaussian distribution, i.e.,

$$\begin{aligned} p(\boldsymbol{\beta}_l|\mathbf{x}_{\text{in},l}; \mathbf{b}, \mathbf{P}, \sigma^2) &= \frac{p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P})p(\mathbf{x}_{\text{in},l}|\boldsymbol{\beta}_l; \sigma^2)}{\int p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P})p(\mathbf{x}_{\text{in},l}|\boldsymbol{\beta}_l; \sigma^2)d\boldsymbol{\beta}_l} \\ &\sim \mathcal{CN}(\boldsymbol{\mu}_l, \boldsymbol{\Xi}) \end{aligned} \tag{19}$$

with

$$\boldsymbol{\mu}_l = \boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}}\left(\mathbf{D}_{\text{in}}\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}} + \sigma^2\mathbf{I}_{N_D}\right)^{-1}\mathbf{x}_{\text{in},l} \tag{20}$$

$$\boldsymbol{\Xi} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}}\left(\mathbf{D}_{\text{in}}\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}} + \sigma^2\mathbf{I}_{N_D}\right)^{-1}\mathbf{D}_{\text{in}}\boldsymbol{\Sigma}. \tag{21}$$

Let $\mathbf{T} = \mathbf{D}_{\text{in}}\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}} + \sigma^2\mathbf{I}_{N_D}$. Given interleaved form sample set $\mathbf{X}_{\text{in}} = [\mathbf{x}_{\text{in},1}, \mathbf{x}_{\text{in},2}, \ldots, \mathbf{x}_{\text{in},L}] \in \mathbb{C}^{N_D \times L}$, the type-II maximum likelihood (ML) estimation of $\mathbf{b}$, $\mathbf{P}$, and σ^2 can be written as follows:

$$\{\hat{\mathbf{b}}, \hat{\mathbf{P}}, \hat{\sigma}^2\} = \arg\min_{\mathbf{b},\mathbf{P},\sigma^2} \mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2) \tag{22}$$

where the cost function $\mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2)$ takes the form of

$$\begin{aligned} \mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2) &= -\sum_{l=1}^{L} \ln p\left(\mathbf{x}_{\text{in},l}; \mathbf{b}, \mathbf{P}, \sigma^2\right) \\ &= -\sum_{l=1}^{L} \ln \int p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P}) p\left(\mathbf{x}_{\text{in},l}|\boldsymbol{\beta}_l; \sigma^2\right) d\boldsymbol{\beta}_l \\ &\propto \ln|\mathbf{T}| + \frac{1}{L}\text{Tr}\left(\mathbf{X}_{\text{in}}^{\text{H}}\mathbf{T}^{-1}\mathbf{X}_{\text{in}}\right). \end{aligned} \tag{23}$$

Below, we present detailed estimation methods for parameters $\mathbf{b}$, $\mathbf{P}$, and σ^2. Specifically, σ^2 and $\mathbf{P}$ are estimated by the expectation maximization (EM) algorithm, while $\mathbf{b}$ is estimated via the fixed-point iteration method.

1) Update of the Noise Power σ^2:

Similar to the approach used in [55], the estimate for σ^2 is obtained by maximizing the conditional expectation $\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}[\ln p(\mathbf{x}_{\text{in},l}|\boldsymbol{\beta}_l; \sigma^2)|\mathbf{x}_{\text{in},l}; \mathbf{b}, \mathbf{P}, \sigma^2]$, then we can derive the updating rule of σ^2, which is

$$\sigma^2 = \frac{\|\mathbf{X}_{\text{in}} - \mathbf{D}_{\text{in}}\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}}\mathbf{T}^{-1}\mathbf{X}_{\text{in}}\|_{\text{F}}^2}{LN_D - L\text{Tr}\left(\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}}\mathbf{T}^{-1}\mathbf{D}_{\text{in}}\right)}. \tag{24}$$

2) Update of the Polarization Covariance Matrix $\mathbf{P}$:

Review (23), we can note that only $p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P})$ is related to $\mathbf{P}$. Thus, by maximizing the conditional expectation $\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}[\ln p(\boldsymbol{\beta}_l; \mathbf{b}, \mathbf{P})|\mathbf{x}_{\text{in},l}; \mathbf{b}, \mathbf{P}, \sigma^2]$, we can obtain the updating rule of $\mathbf{P}$, i.e.,

$$\mathbf{P} = \frac{1}{LK}\sum_{l=1}^{L}\sum_{k=1}^{K} b_k^{-1}\left(\boldsymbol{\mu}_l^k\left(\boldsymbol{\mu}_l^k\right)^{\text{H}} + \boldsymbol{\Xi}_k\right) \tag{25}$$

where

$$\boldsymbol{\mu}_l^k = \boldsymbol{\mu}_{l,[N_p(k-1)+1:N_pk]} \tag{26}$$

$$\boldsymbol{\Xi}_k = \boldsymbol{\Xi}_{[N_p(k-1)+1:N_pk, N_p(k-1)+1:N_pk]}. \tag{27}$$

For detailed derivation, see Appendix B. ■

However, the updating rule of $\mathbf{P}$ in (25) will bring huge computational complexity which is not conducive to real-time processing. Considering that CNR is relatively high in the actual clutter environment, we have the following approximation:

$$\mathbf{T}^{-1} \approx \left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}} + \sigma^2\mathbf{I}_{NM}\right)^{-1} \otimes \mathbf{P}^{-1}. \tag{28}$$

For detailed discussion of (28), see Appendix C. ■

It is important to note that the relaxation in (28) is only to reduce the computational complexity of estimating $\mathbf{P}$, which is different from the Kronecker structured estimators of reconstructing CNCM. Thereby, $\boldsymbol{\Xi}$ in (21) can be simplified in a manner similar to [56], that is,

$$\boldsymbol{\Xi}_k \approx p_c^k\mathbf{P},\ k = 1, 2, \ldots, K \tag{29}$$

where p_c^k is the kth diagonal element of the posterior covariance matrix [similar to (21)] of the clutter space-time profile, and (25) can be rewritten as follows:

$$\mathbf{P} = \frac{1}{LK - \sum_{l=1}^{L}\sum_{k=1}^{K} p_c^k b_k^{-1}}\sum_{l=1}^{L}\sum_{k=1}^{K} b_k^{-1}\boldsymbol{\mu}_l^k\left(\boldsymbol{\mu}_l^k\right)^{\text{H}}. \tag{30}$$

Importantly, we do not need to calculate the specific value of p_c^k, since the normalization process $\mathbf{P} = \mathbf{P}/\|\mathbf{P}\|_{\mathrm{F}}$ is employed to remove the ambiguity between the polarization domain and the space-time domain. At this point, we can achieve the simplified updating rule for $\mathbf{P}$ with lower computational complexity, that is,

$$\mathbf{P} = \sum_{l=1}^{L}\sum_{k=1}^{K} b_k^{-1}\boldsymbol{\mu}_l^k\left(\boldsymbol{\mu}_l^k\right)^{\mathrm{H}}, \quad \mathbf{P} = \frac{\mathbf{P}}{\|\mathbf{P}\|_{\mathrm{F}}}. \tag{31}$$

3) Update of the Clutter Space-Time Power $\mathbf{b}$*:*

For the estimate updating rule of $\mathbf{b}$, the abovementioned EM algorithm generally converges slowly. Herein, we exploit the block diagonal structure of $\boldsymbol{\Sigma}$, and take derivatives of $\mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2)$ with respect to (w.r.t) b_k, $k = 1, 2, \ldots, K$. First, using the fact that

$$\frac{\partial \ln|\mathbf{T}|}{\partial b_k} = \mathrm{Tr}\left(\mathbf{T}^{-1}\frac{\partial \mathbf{T}}{\partial b_k}\right), \quad \frac{\partial \mathbf{T}^{-1}}{\partial b_k} = -\mathbf{T}^{-1}\frac{\partial \mathbf{T}}{\partial b_k}\mathbf{T}^{-1} \tag{32}$$

$$\frac{\partial \mathbf{T}}{\partial b_k} = \mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}} \tag{33}$$

where $\mathbf{D}_{\mathrm{in},k} = \mathbf{D}_{\mathrm{in}}[:, N_p(k-1)+1 : N_p k]$. Then, differentiate $\mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2)$ w.r.t b_k, yielding

$$\begin{aligned}\frac{\partial \mathcal{L}\left(\mathbf{b}, \mathbf{P}, \sigma^2\right)}{\partial b_k} &= \frac{\partial\left(\ln|\mathbf{T}| + \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_{\mathrm{in},l}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{x}_{\mathrm{in},l}\right)}{\partial b_k} \\ &\propto \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_{\mathrm{in},l}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{x}_{\mathrm{in},l} \\ &\quad - \mathrm{Tr}\left(\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\right).\end{aligned} \tag{34}$$

Set $\partial\mathcal{L}(\mathbf{b}, \mathbf{P}, \sigma^2)/\partial b_k = 0$, having

$$\frac{1}{L}\frac{\sum_{l=1}^{L}\mathbf{x}_{\mathrm{in},l}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{x}_{\mathrm{in},l}}{\mathrm{Tr}\left(\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\right)} = 1 = \frac{b_k}{b_k}. \tag{35}$$

Then, we modify (35) by introducing the factor $(b_k^{(t)}/b_k^{(t-1)})$ to obtain the fixed-point updating rule for b_k, i.e.,

$$\frac{b_k^{(t)}}{b_k^{(t-1)}} = \frac{\sum_{l=1}^{L}\mathbf{x}_{\mathrm{in},l}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\mathbf{T}^{-1}\mathbf{x}_{\mathrm{in},l}}{L\,\mathrm{Tr}\left(\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\right)} \tag{36}$$

where the superscript (t) means the result of the (t)th iteration. However, the convergence speed of this version is fixed. To address this, we introduce an exponent term q to flexibly control the convergence speed of $\mathbf{b}$. The ultimate updating rule evolves to

$$b_k^{(t)} = b_k^{(t-1)}\left(\frac{\mathrm{Tr}(\mathbf{R}_{\mathrm{in,SCM}}\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}}\mathbf{T}^{-1})}{\mathrm{Tr}(\mathbf{T}^{-1}\mathbf{D}_{\mathrm{in},k}\mathbf{P}\mathbf{D}_{\mathrm{in},k}^{\mathrm{H}})}\right)^{q} \tag{37}$$

where $\mathbf{R}_{\mathrm{in,SCM}} = \mathbf{X}_{\mathrm{in}}\mathbf{X}_{\mathrm{in}}^{\mathrm{H}}/L$.

Remark 2: By introducing the exponent term q, one can flexibly control the step size from $b_k^{(t-1)}$ to $b_k^{(t)}$, thereby controlling the convergence speed of SAW-SBL PSTAP. The larger q is, the faster the algorithm converges, but the parameter estimation accuracy deteriorates or even diverges, affecting the clutter suppression performance; the smaller q is, although the parameter estimation accuracy and clutter suppression performance improved, the algorithm converges slowly and the computational complexity increases significantly. Section IV also demonstrates the above analysis. For practical applications, we can set the initial value $q = 1$ and balance output SCNR and real-time performance according to the scenario requirements to decide whether to reduce or increase q.

Algorithm 1 SAW-SBL PSTAP

Input: $\mathbf{D}$, $\mathbf{D}_{in}$, $\{\mathbf{x}_{in,l}\}_{l=1}^{L}$, q, $t_{\max}$, and μ

1: **Initialize:** $b_k = 1\ \forall k$, $\mathbf{P} = \mathbf{I}_{N_p}$, and $\sigma^2 = 10$
2: **Repeat**
3: Update $\left(\sigma^2\right)^{(t)}$ via (24)
4: Update $\mathbf{P}^{(t)}$ via (31)
5: Update $b_k^{(t)}$ via (37)
6: Compute $\boldsymbol{\Sigma}^{(t)} = \mathrm{diag}(\mathbf{b}^{(t)}) \otimes \mathbf{P}^{(t)}$
7: **Until** $t = t_{\max}$ or $\|\boldsymbol{\Sigma}^{(t)} - \boldsymbol{\Sigma}^{(t-1)}\|_{\mathrm{F}}/\|\boldsymbol{\Sigma}^{(t)}\|_{\mathrm{F}} \le \mu$
8: Compute $\hat{\mathbf{R}}_{in}$ via (38) or $\hat{\mathbf{R}}$ via (39)

Output: $\hat{\mathbf{R}}_{in}$ or $\hat{\mathbf{R}}$

Ultimately, after obtaining the estimation of $\mathbf{b}$, $\mathbf{P}$, and σ^2, the CNCM can be reconstructed via

$$\hat{\mathbf{R}}_{\mathrm{in}} = \mathbf{D}_{\mathrm{in}}(\mathrm{diag}(\hat{\mathbf{b}}) \otimes \hat{\mathbf{P}})\mathbf{D}_{\mathrm{in}}^{\mathrm{H}} + \hat{\sigma}^2\mathbf{I}_{N_D} \tag{38}$$

$$\hat{\mathbf{R}} = \mathbf{D}(\hat{\mathbf{P}} \otimes \mathrm{diag}(\hat{\mathbf{b}}))\mathbf{D}^{\mathrm{H}} + \hat{\sigma}^2\mathbf{I}_{N_D}. \tag{39}$$

The proposed SAW-SBL algorithm for PSTAP is summarized in Algorithm 1, where $t_{\max}$ denotes the maximum number of iterations and μ is the preset convergence threshold.

C. Computational Complexity

In SR-based STAP algorithms, the computational complexity mainly from the CNCM estimation, including the iterative solutions of $\mathbf{b}$ and σ^2. For PSTAP, on one hand, the dimension of the received data has enlarged due to different polarization channels being introduced, on the other hand, in addition to $\mathbf{b}$ and σ^2, the proposed method also adds an iterative solution for $\mathbf{P}$, thus the computational complexity increases. Specifically, note the Kronecker structure of $\mathbf{D}_{\mathrm{in}}$ and the block structure of $\boldsymbol{\Sigma}$. For each iteration, the computational complexity of $\mathbf{T}^{-1}$ is $\mathcal{O}(N_D^3 + N_D^2 N_K + N_p N_D N_K)$, the computational complexity of $\mathbf{P}$ in (25) is $\mathcal{O}(N_D N_K L(N_D+1) + N_D N_K N_p)$, while the computational complexity of $\mathbf{P}$ in (31) is reduced to $\mathcal{O}(N_D N_K L(N_D+1))$, the computational complexity of σ^2 and $\mathbf{b}$ is $\mathcal{O}(N_D^3 + N_D^2(N_K + L) + N_D(N_K L + N_K N_p + L))$. Denote that the number of iterations is N_t and the overall computational complexity of SAW-SBL PSTAP is $\mathcal{O}(N_t(2N_D^3 + N_D^2(N_K L + 2N_K + L) + N_D(2N_K L + 2N_K N_p + L)))$.

For comparison, considering that the fast converging SBL (FCSBL) algorithm [49] performs well in convergence speed and clutter cancellation, we take FCSBL-D PSTAP [as defined in (12)] and FCSBL-K PSTAP [as defined in (14)] into account. Notably, FCSBL-D does not estimate $\mathbf{P}$, while FCSBL-K only leverages FCSBL to estimate the clutter space-time profiles and noise power, reconstructing the CNCM by (14). Thereby, the overall computational complexity of FCSBL-D and FCSBL-K are $\mathcal{O}(N_t(2N_D^3 + N_D^2(2N_K + L) + N_D(N_K L + N_K + L)))$ and $\mathcal{O}(N_t(2M^3N^3 + M^2N^2(2K+L) + MN(KL+K+L)))$, respectively. In addition, lower complexity algorithms such as the loaded SCM (LSCM) and robust

shrinkage Kronecker estimator (RSKE) [33], with complexities of $\mathcal{O}(LN_D^2)$ and $\mathcal{O}(N_t(M^3N^3+N_p^3+M^2N^2(LN_p+1)+N_p^2(LMN+1)))$, can also be applied to PSTAP. Although SAW-SBL yields higher computational complexity, Section IV demonstrates its superior low-speed target detection performance in heterogeneous clutter environments.

IV. Numerical Experiments

In the following, we evaluate the performance of clutter suppression and target detection through simulated and measured clutter data, respectively. Considering that the FCSBL algorithm [49] performs well in convergence speed and clutter cancellation, SAW-SBL PSTAP is compared with FCSBL-D PSTAP [as shown in (12)] and FCSBL-K PSTAP [as shown in (14)]. Meanwhile, we also take the classical LSCM (the loading factor is set to the noise level) and statistical framework-based Kronecker estimator RSKE [33] into account.

Unless otherwise stated, we focus on small sample cases, $L=6$ is for the simulated clutter data, while $L=4$ is for the measured clutter data. The dictionary size for SR-based methods is selected as $K=36MN$, and the initialization values throughout the experiments are set as follows: $\eta=1$, $b_k=1\ \forall k$, $\mathbf{P}=\mathbf{I}_{N_p}$, $\sigma^2=10$, $t_{\max}=50$, and $\mu=10^{-3}$.

A. Simulated Data

In this section, we demonstrate the effectiveness of the proposed method by several simulation experiments. The simulated clutter data is generated from the model mentioned in Section II-A, and partial simulation parameters are listed as follows: $N=6$ dual-polarized ULA antennas with half wavelength interchannel spacing receive $M=6$ pulses in a CPI, PRF is 2000 Hz, wavelength is 0.25 m, bandwidth is 5 MHz, the platform height is 6 km, and the platform velocity is 125 m/s. The CNR is 40 dB, clutter power ratio is $\chi_c=1$, clutter statistical average phase difference is $\bar{\eta}_c=\pi/2$, and each range cell contains $N_c=181$ clutter patches. The target power ratio is $\chi_t=1$, target statistical average phase difference is $\bar{\eta}_t=7\pi/6$, and target correlation coefficient is $\rho_t=1$ (considered as a completely polarized wave). Based on these conditions, the polarization difference between target and clutter is relatively large, and the detection performance for low-speed targets of PSTAP will improve as $|\rho_c|$ increases [14]. The output SCNR loss is adopted to assess the clutter suppression performance of different methods, having

$$\text{SCNR loss}=\frac{\sigma^2}{N_pNM}\frac{|\hat{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\text{pst}}^t|^2}{\hat{\mathbf{w}}^{\mathrm{H}}\mathbf{R}\hat{\mathbf{w}}} \tag{40}$$

where $\hat{\mathbf{w}}$ denotes the estimated PSTAP weight vector, and all results are averages of 1000 independent Monte-Carlo trials.

First, we assess the complexity of SAW-SBL. Table I compares the global and single iteration runtimes of SAW-SBL, FCSBL-K, FCSBL-D, and RSKE, all using the same software (MATLAB R2021b) and hardware configuration: CPU: Intel[1] Core[2] i7-10700@2.90 GHz; RAM: 16 GB. Since SAW-SBL needs to estimate more parameters using the iterative solution framework, its runtime is longer than that of the other three algorithms. The runtime of FCSBL-K is faster than that of FCSBL-D but longer than that of RSKE, and the estimated runtimes corroborate the theoretical analysis. The reported runtimes are specific to the described software and hardware platform. In real-time applications, the runtime can be reduced by using high-performance hardware and parallel computing. Moreover, developing a tensor-based [50] SAW-SBL framework can significantly reduce computational complexity by transforming large-scale dictionary operations into lower complexity tensor representations, effectively mitigating redundant data operations, high-dimensional matrix multiplications, and inversions. Fig. 3 presents the required memory (here approximated by the number of variables) of different algorithms versus MN, we can see that as MN increases, more memory/variables are needed (except for RSKE, since RSKE only solves one variable, the shrinkage factor, which is independent of the size of MN, while the number of variables of the other three algorithms depends on the number of grid points K and is related to MN). SAW-SBL exceeds FCSBL-K and RSKE but is comparable to FCSBL-K and lower than FCSBL-D. For practical application scenarios, e.g., larger datasets or nonideal clutter conditions, a larger K is typically required to maintain effective clutter suppression performance, this will lead to increased runtime and memory usage.

TABLE I
Iteration Runtime Comparison of Different Algorithms

Algorithm	Global runtime (s)	Single iteration runtime (s)
SAW-SBL	3.0701	0.2790
FCSBL-K	0.0717	0.0080
FCSBL-D	0.3455	0.0384
RSKE	0.0074	0.0012

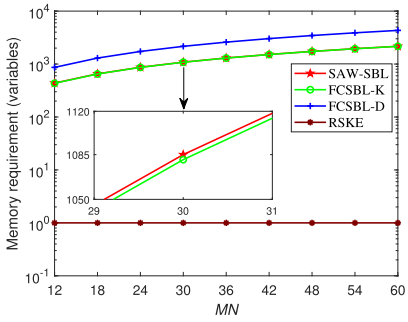


Fig. 3. Memory usage comparison of different algorithms versus *MN*.

Next, we employ the cost function $\ln|\hat{\mathbf{R}}|+\mathrm{Tr}(\hat{\mathbf{R}}^{-1}\mathbf{R}_{\text{SCM}})$ [49] to assess the convergence performance of SBL-based algorithms with different clutter complex correlation coefficients ρ_c. In Fig. 4(a), if $|\rho_c|=0.1$, the polarization covariance matrix $\mathbf{R}_p$ is approximately a diagonal matrix, as the number of iterations increases, SAW-SBL and FCSBL-D almost converge to the same value. Since FCSBL-K only estimates partial parameters using SR techniques, its ultimate value of cost function is slightly higher than that of SAW-SBL and FCSBL-D. Meanwhile, we can see that SAW-SBL can flexibly control the convergence speed by choosing different q,

[1]Registered trademark.
[2]Trademarked.

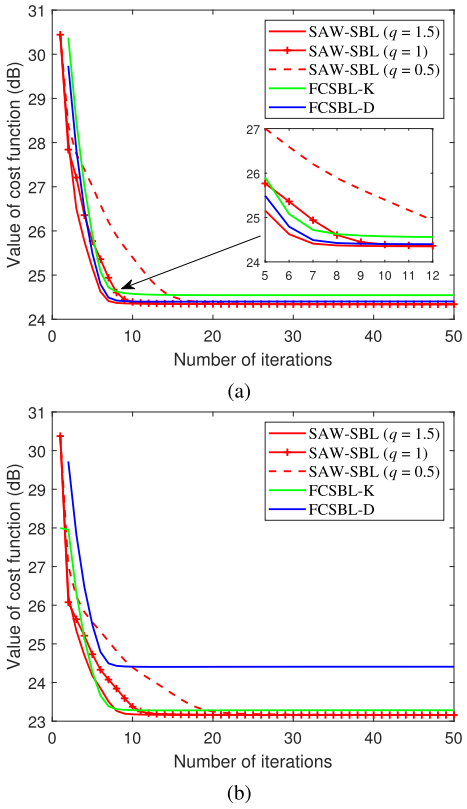


Fig. 4. Value of cost function versus the number of iterations. (a) $|\rho_c| = 0.1$. (b) $|\rho_c| = 0.99$.

the convergence speed increases as q increases. When $q = 1.5$, the convergence speed is faster than FCSBL-based methods. In Fig. 4(b), if $|\rho_c| = 0.99$, $\mathbf{R}_p$ is obviously not a diagonal matrix, the ultimate cost function value of FCSBL-D is higher than FCSBL-K and SAW-SBL due to the structure loss as mentioned in Section III-A. FCSBL-K avoids this structural loss, but its ultimate cost-function value is still higher than that of SAW-SBL.

To proceed, it is important to note that we can control the convergence speed of SAW-SBL through q, but this will also affect the clutter suppression performance. Fig. 5(a) gives the impact of different q on SCNR loss with $|\rho_c| = 0.99$, we can see that as q increases, the clutter suppression performance decreases accordingly. The simulation results presented above validate the analysis of SAW-SBL in Section III-B, to balance convergence speed and clutter suppression performance, we set $q = 1$ in the subsequent experiments. It is also important to note that η affects the performance of FCSBL-K. To verify it, we consider two intrinsic clutter motion situations (i.e., the velocity standard deviations are 0.05 and 2, respectively) to represent different clutter environments. Fig. 5(b) and (c) shows the impact of different η on SCNR loss, we can see that as η increases, the detection performance of low-speed targets deteriorates. When η is too small, the overall clutter suppression performance deteriorates. Meantime, by comparing Fig. 5(b) and (c), the clutter suppression performance of setting $\eta = 1.5$ and $\eta = 1$ is opposite in different clutter environments, which indicates that fixed η is not suitable for time-varying clutter environments. The above simulation results verify the analysis of FCSBL-K in Section III-A. To achieve satisfactory target detection performance, we set $\eta = 1$ for FCSBL-K in subsequent experiments.

Second, we first exhibit the SCNR loss curves of various methods with different $|\rho_c|$ in the ideal case. Fig. 6 presents the results in the ideal case, as shown in Fig. 6(a), when $|\rho_c| = 0.1$, SBL-based methods achieve almost optimal performance (compared to the black curve). RSKE uses the Kronecker structure and adaptively selects the shrinkage factor to achieve better clutter suppression performance. Although it is more suitable for small sample cases than LSCM, it still performs poorly compared with SBL-based methods. Since the clutter correlation is low in this case, all algorithms do not show the detection advantage of PSTAP for low-speed targets.

When the clutter correlation becomes higher, i.e., $|\rho_c| = 0.99$, as shown in Fig. 6(b), SAW-SBL and FCSBL-K exhibit the detection advantage of PSTAP for low-speed targets, and SAW-SBL performs well than FCSBL-K. Due to the structure loss, FCSBL-D does not promote the detection performance of low-speed targets. Although RSKE takes the choice of shrinkage operators into account, small sample sizes reduce the estimation accuracy of $\mathbf{R}_p$, leading to performance degradation, even worse than LSCM in this case. Compared to Fig. 6(a), the performance of LSCM has improved because the clutter DoFs decreased in the large correlation case (as illustrated in Section II-B).

To further illustrate the effect of the CNCM reconstruction approach on clutter suppression performance, based on Fig. 6(b), we increase the number of samples to $L = 2N_D$. Fig. 7 provides the SCNR loss curves of various methods with $|\rho_c| = 0.99$ and $L = 2N_D$ in the ideal case. As we can see, even if the number of samples is sufficient, FCSBL-D and RSKE are still unable to detect low-speed targets due to the structural loss of reconstructing CNCM, verifying the analysis in Section III-A and the proof in Appendix A. The overall performance of the other three estimators has improved, and SAW-SBL performs better in this case.

In actual scenarios, inevitable nonideal factors, such as array gain-phase error and intrinsic clutter motion [57], will lead to model mismatch that degrades the performance of SR-based algorithms. Additionally, range ambiguity [58] will lead to a reduction in clutter polarization correlation. Fig. 8 comprehensively considers the above factors and presents the SCNR loss curves when $|\rho_c| = 0.5$ in the nonideal case. We can see that the clutter suppression performance of SBL-based methods has declined but is still better than RSKE and LSCM. SAW-SBL performs better than FCSBL-K and FCSBL-D, and the gap between SAW-SBL and FCSBL-K has increased because the latter is not robust in this scenario.

Third, to evaluate the impact of clutter correlation $|\rho_c|$ on the low-speed target detection performance, Fig. 9 gives the result of average SCNR (aSCNR) loss 1 (the mean value of SCNR loss within normalized Doppler frequency $[-0.025, 0.025]$) versus $|\rho_c|$ in the ideal case. In Fig. 9, as the correlation coefficient $|\rho_c|$ increases, the detection performance of SAW-SBL, FCSBL-K, and LSCM for low-speed targets improves, but this does a slight impact on FCSBL-D and even has a negative impact on RSKE.

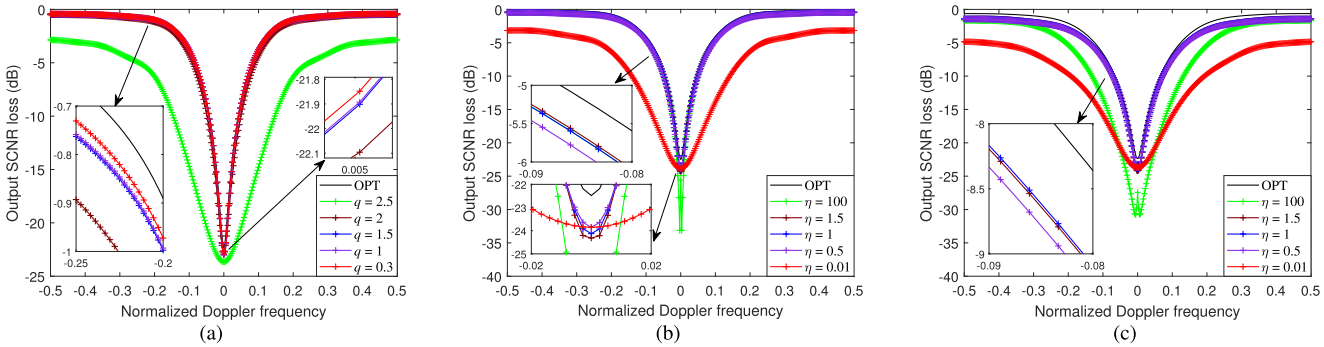


Fig. 5. (a) Impact of q on SCNR loss of SAW-SBL PSTAP with $|\rho_c| = 0.99$. (b) Impact of η on SCNR loss of FCSBL-K PSTAP with $|\rho_c| = 0.99$ and clutter velocity standard deviation 0.05. (c) Impact of η on SCNR loss of FCSBL-K PSTAP with $|\rho_c| = 0.99$ and clutter velocity standard deviation 2.

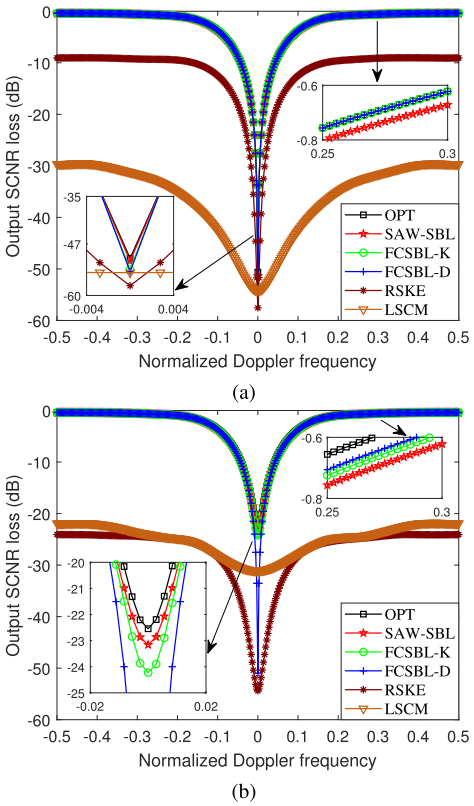


Fig. 6. SCNR loss versus normalized Doppler frequency in the ideal case. (a) $|\rho_c| = 0.1$. (b) $|\rho_c| = 0.99$.

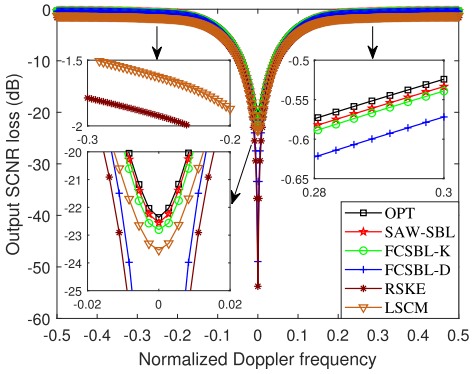


Fig. 7. SCNR loss versus normalized Doppler frequency in the ideal case when $|\rho_c| = 0.99$ and $L = 2N_D$.

Furthermore, to evaluate the impact of sample size on the clutter suppression performance, Fig. 10 depicts the aSCNR loss 2 (the mean value of SCNR loss within the entire normalized Doppler frequency range) curves with different $|\rho_c|$ in the ideal case. From Fig. 10(a), when $|\rho_c| = 0.1$, SBL-based methods converge quickly as the number of training samples increases and can achieve near-optimal performance with a few number of samples. In this case, the improvement of SAW-SBL in low-speed target detection is not obvious, but its detection performance for the high-speed target is slightly worse than FCSBL-K and FCSBL-D [as shown in Fig. 6(a)], so the overall performance is reduced and more samples are required to achieve near-optimal performance. Although the convergence speed of RSKE is faster than that of LSCM, it is much slower than that of SBL-based methods.

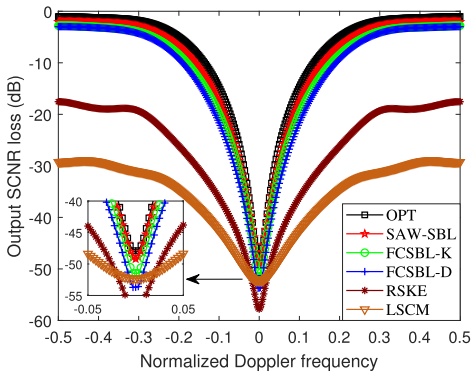


Fig. 8. SCNR loss versus normalized Doppler frequency in the non-ideal case with channel error (0.03,π/90) and clutter velocity standard deviation 2 when $|\rho_c| = 0.5$.

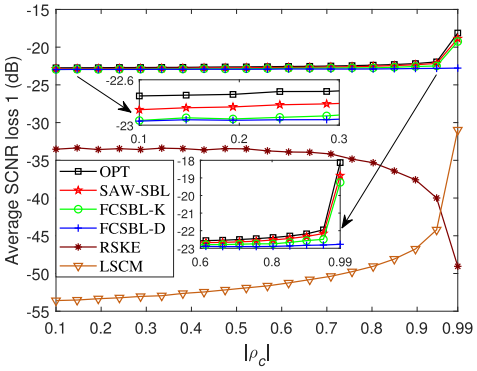


Fig. 9. aSCNR loss 1 versus $|\rho_c|$.

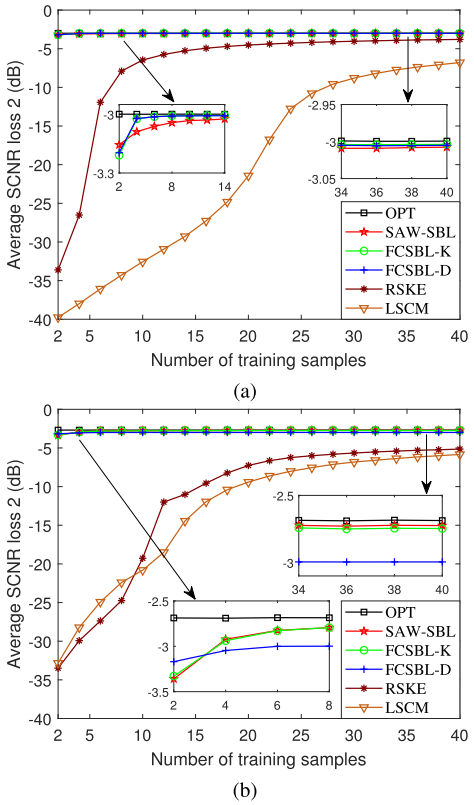


Fig. 10. aSCNR loss 2 versus the number of training samples. (a) $|\rho_c| = 0.1$. (b) $|\rho_c| = 0.99$.

As shown in Fig. 10(b), when $|\rho_c| = 0.99$, although the convergence speed of FCSBL-D is fastest, its final performance is less than that of SAW-SBL and FCSBL-K. Since SAW-SBL performs well for low-speed target detection in this case, the overall performance has improved. Its convergence speed is almost the same as FCSBL-K, but the final performance exceeds FCSBL-K as the number of samples increases. Note that SAW-SBL is more robust than FCSBL-K, as illustrated in Fig. 8. Although the convergence speed of RSKE is faster than that of LSCM, RSKE performs worse than LSCM in this case when the sample is limited [can also be noted in Fig. 6(b)]. Meantime, we can see that as the number of samples increases, the aSCNR loss 2 of RSKE in Fig. 10(b) converges to a lower value than that in Fig. 10(a), this indicates that the increased clutter polarization correlation does not enhance RSKE's detection performance for low-speed targets.

B. Measured Data

In this section, we examine the target detection performance of the proposed algorithm with the measured PSTAP data. The measured data is recorded by an airborne dual-polarized radar system, one can refer to [59] for a more detailed analysis of the measured data. The radar system deploys a ULA of 32 diversely polarized antennas, working in the side-looking mode. Its main beam points to $\theta_0 = 0°$ and $\varphi_0 = 85°$, the V channel transmits signals, and both H and V channels receive signals. In the dataset, 314 pulses are collected in one CPI, containing a total of 643 range snapshots, and all the data have been pulse-compressed. The Capon spectrums of both the

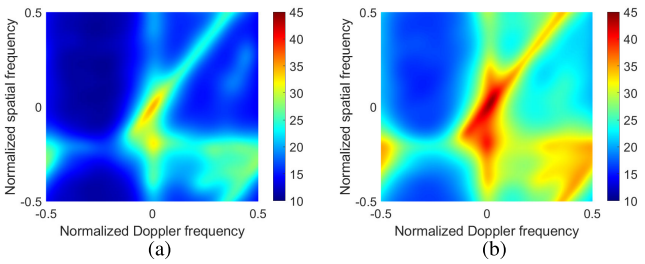


Fig. 11. Capon spectrum of the measured clutter data. (a) H channel clutter. (b) V channel clutter.

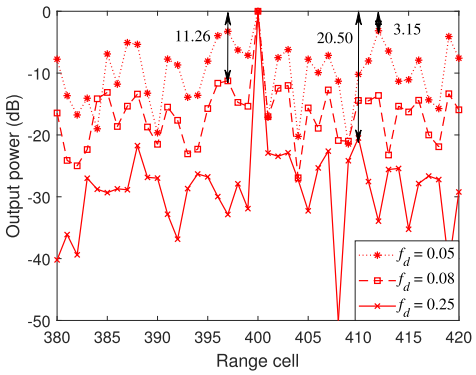


Fig. 12. Detection performance of the proposed SAW-SBL algorithm for targets with different velocities.

H and V channel clutter estimated by all 643 snapshots are depicted in Fig. 11(a) and (b), respectively. As we can see, the clutter power of V channel is much stronger than H channel, this is related to the V channel transmitting signal.

The target located on the mainlobe with polarization vector $[0.9992, -0.0398 - 0.0052i]^{\mathrm{T}}$ (there is a relatively large polarization distance between clutter and target) is artificially added in the 400th range cell. Based on the clutter power calculated in the H channel, the signal-to-clutter ratio (SCR) is set to -5 dB. We consider the target with different normalized Doppler frequency f_d, i.e., $f_d = 0.05, 0.08, 0.25$, respectively. Meantime, to verify the detection performance of aforesaid PSTAP algorithms for the low-speed target, we also take the FCSBL-STAP method to separately detect H channel target and V channel target, denoted as FCSBL-H and FCSBL-V. Since the adaptive matched filter (AMF) [3] owns the constant false alarm rate property, we employ AMF to detect targets

$$P_{\mathrm{AMF}} = \frac{\left|\mathbf{a}^{\mathrm{H}}\hat{\mathbf{R}}^{-1}\mathbf{x}\right|^2}{\mathbf{a}^{\mathrm{H}}\hat{\mathbf{R}}^{-1}\mathbf{a}}. \tag{41}$$

Recall that $\mathbf{a}$ denotes the polarization-space-time or space-time steering vector of the desired signal, $\hat{\mathbf{R}}$ denotes the estimated CNCM by above methods, and $\mathbf{x}$ denotes the received dual-polarized or single-polarized echo data.

The processing window includes 41 cells, from the 380th range cell to the 420th range cell. For each CUT, four samples symmetrically positioned around it are regarded as the guard cells. Employing the spectral similarity method in [60], we select four samples from 20 candidate samples to estimate the CNCM. In light of computational considerations, the subaperture data derived from six consecutive sensors and

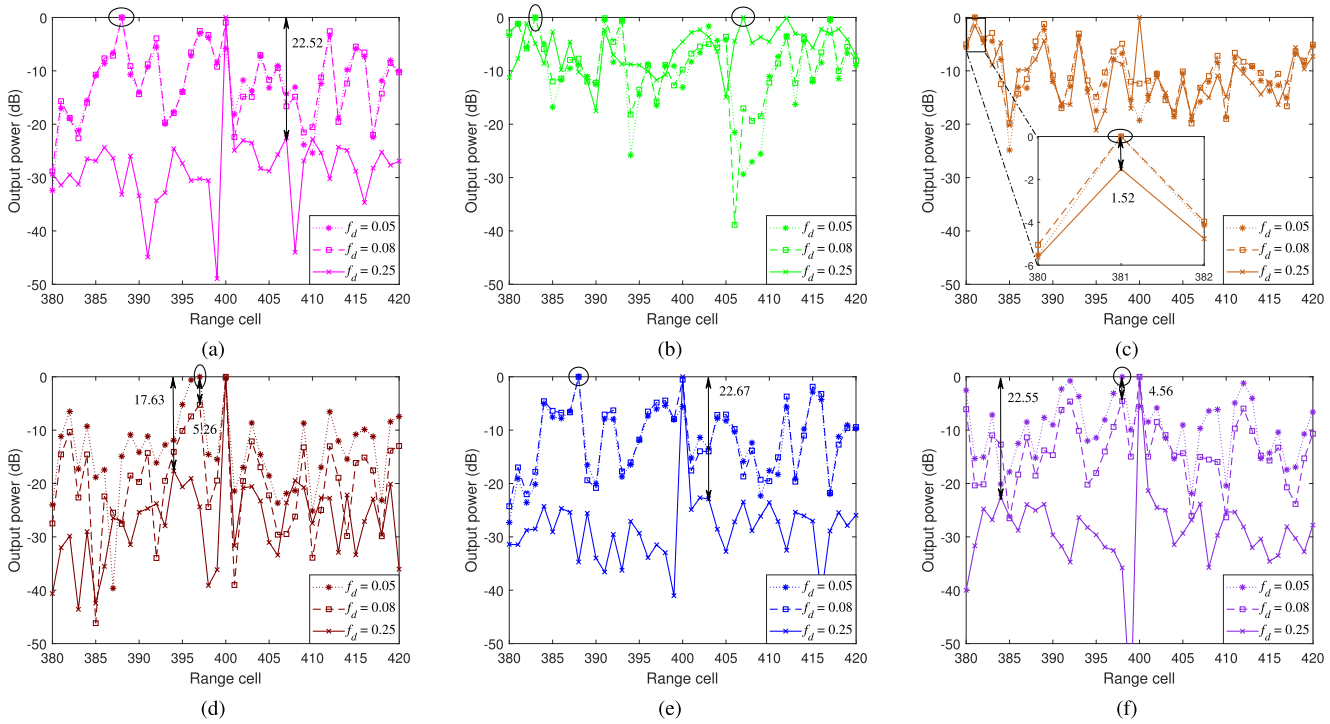


Fig. 13. Detection performance of various comparison algorithms for targets with different velocities. (a) FCSBL-H. (b) FCSBL-V. (c) LSCM. (d) RSKE. (e) FCSBL-D. (f) FCSBL-K.

TABLE II
TARGET DETECTION PERFORMANCE OF DIFFERENT ALGORITHMS

Method	$f_d = 0.25$			$f_d = 0.08$			$f_d = 0.05$		
	Target detected	PSL	aSCNR	Target detected	PSL	aSCNR	Target detected	PSL	aSCNR
FCSBL-H	✓	22.52 dB	28.47 dB	×			×		
FCSBL-V	×			×			×		
LSCM	✓	1.52 dB	11.42 dB	×			×		
RSKE	✓	17.63 dB	26.98 dB	✓	5.26 dB	20.20 dB	×		
FCSBL-D	✓	22.67 dB	28.28 dB	×			×		
FCSBL-K	✓	22.55 dB	28.93 dB	✓	4.56 dB	14.34 dB	×		
SAW-SBL	✓	20.50 dB	28.79 dB	✓	11.26 dB	16.90 dB	✓	3.15 dB	10.20 dB

six pulses are applied in subsequent experiments. Utilizing the sliding window approach, Figs. 12 and 13 present the normalized AMF output of the proposed SAW-SBL method and various comparison algorithms, respectively. Similar to [61], if the target can be detected, i.e., the AMF output power of the range cell where the target is located is the largest among all range cells. We use the peak sidelobe level (PSL) as one of the evaluation metrics, it is calculated by the difference between the output power of the target-detected range cell and the second-largest output power within all range cells except guard cells, and the performance on target detection is better if the PSL is larger. If the target is not detected, the maximum AMF output power is not the range cell where the target is located, marked within the circle in the subplot.

As shown in Figs. 12 and 13, for the high-speed target, i.e., $f_d = 0.25$, since the clutter power of the V channel is much higher than that of the H channel (explained in Fig. 11), causing a lower SCR, all algorithms except FCSBL-V have achieved satisfactory target detection performance. SBL-based algorithms can own better target detection performance than LSCM at low sample support. When the target velocity decreases, that is, $f_d = 0.08$, the STAP algorithm, FCSBL-H, can no longer detect the target correctly. For PSTAP algorithms, the target is not detected by LSCM due to a lack of training samples. Affected by the loss of CNCM structure, FCSBL-D also can not detect the target. SAW-SBL, FCSBL-K, and RSKE can detect the target in this case, among which SAW-SBL has the largest PSL. When the target speed further decreases, that is, $f_d = 0.05$, only SAW-SBL can detect the target, and the PSL is 3.15 dB. This indicates that SAW-SBL can give full play to the detection performance of PSTAP for the low-speed target under limited samples.

Meantime, we also employ the aSCNR as another evaluation metric, it is calculated by the difference between the output power of the target-detected range cell and the average value of the output power of all range cells except both the target-detected range cell and guard cells. For clarity, we summarize the total results in Table II. Note that model mismatch will affect the performance of SBL-based methods, but how to mitigate the impact of model mismatch is not the

focus of this article. Denser dictionary [50] and dictionary learning [62]-based methods have been presented to overcome this issue arising from model mismatch, by leveraging these advanced technologies, the performance of SAW-SBL can be further improved. All the results demonstrate that the proposed SAW-SBL method can achieve superior target detection performance, especially for low-speed targets.

V. Conclusion

In this article, we present the SAW-SBL method for PSTAP in heterogeneous clutter environments. We theoretically analyze the limited performance of Kronecker estimators for low-speed target detection. With finite samples, directly introducing traditional SR-STAP methods into PSTAP will lead to the loss of CNCM structure, which also affects the detection performance of low-speed targets. To accurately reconstruct the CNCM and ensure its matrix structure, we exploit both the Kronecker characteristic and the intrinsic sparsity of clutter, then formulate the type-II ML estimation as a block SR problem. We derive a fast converging and controllable learning framework to iteratively update the noise power, polarization covariance matrix, and clutter space-time power, resulting in accurate CNCM estimation. Extensive experimental results verify the effectiveness of SAW-SBL PSTAP, particularly its superior performance in detecting low-speed targets.

For future work, we aim to further reduce the computational complexity and enhance the robustness of SAW-SBL PSTAP in the presence of model mismatch.

Appendix A
Proof of Proposition 1

Recall that the Kronecker estimator can be formulated as follows:

$$\hat{\mathbf{R}}_{\text{Kro}} = \hat{\mathbf{R}}_p \otimes \hat{\mathbf{R}}_{\text{st}}. \tag{42}$$

Leveraging the Kronecker product property $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ and $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC}\otimes\mathbf{BD})$, the PSTAP filter weight vector can be calculated via (7), i.e.,

$$\begin{aligned}\mathbf{w}_{\text{Kro}} &= \tau_{\text{Kro}}\hat{\mathbf{R}}_{\text{Kro}}^{-1}\mathbf{a}_{\text{pst}}^{t}\\ &= \tau_{\text{Kro}}\left(\hat{\mathbf{R}}_p^{-1}\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\right)\left(\mathbf{p}_t\otimes\mathbf{a}_{\text{st}}^{t}\right)\\ &= \tau_{\text{Kro}}\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\end{aligned} \tag{43}$$

where $\tau_{\text{Kro}} = \left(\mathbf{a}_{\text{pst}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{Kro}}^{-1}\mathbf{a}_{\text{pst}}^{t}$. Using the Kronecker product property $(\mathbf{A}\otimes\mathbf{B})^{\text{H}} = \mathbf{A}^{\text{H}}\otimes\mathbf{B}^{\text{H}}$, the output SCNR of Kronecker estimators is

$$\begin{aligned}\text{SCNR}_{\text{Kro}} &= \frac{|\alpha_t|^2\left|\mathbf{w}_{\text{Kro}}^{\text{H}}\mathbf{a}_{\text{pst}}^{t}\right|^2}{\mathbf{w}_{\text{Kro}}^{\text{H}}\mathbf{R}\mathbf{w}_{\text{Kro}}}\\ &= \frac{|\alpha_t|^2\left(\left(\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\left(\mathbf{p}_t\otimes\mathbf{a}_{\text{st}}^{t}\right)\right)^2}{\left(\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\mathbf{R}\left(\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\right)}\\ &= \frac{|\alpha_t|^2\left(\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\right)^2}{\left(\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-1}\otimes\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-1}\right)\mathbf{R}\left(\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\otimes\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\right)}.\end{aligned} \tag{44}$$

For notational simplicity, let $|\alpha_t|^2 = 1$, $\hat{\eta}_p = \mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t$ and $\hat{\eta}_{\text{st}} = \left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}$. Substituting the ideal CNCM $\mathbf{R} = \mathbf{R}_p\otimes\mathbf{R}_{\text{st}} + \sigma^2\mathbf{I}_{N_D}$ into (44), then (44) can be simplified to

$$\text{SCNR}_{\text{Kro}} \triangleq \frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_c} \tag{45}$$

where $\hat{\eta}_c$ is the output clutter plus noise power, having

$$\begin{aligned}\hat{\eta}_c = {}& \mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-1}\mathbf{R}_p^{-1}\hat{\mathbf{R}}_p^{-1}\mathbf{p}_t\cdot\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{R}_{\text{st}}^{-1}\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t}\\ &+\sigma^2\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-2}\mathbf{p}_t\cdot\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-2}\mathbf{a}_{\text{st}}^{t}.\end{aligned} \tag{46}$$

As the number of samples increases, $\lim_{l\to+\infty}\hat{\mathbf{R}}_p \approx \mathbf{R}_p$, $\lim_{l\to+\infty}\hat{\mathbf{R}}_{\text{st}} \approx \mathbf{R}_{\text{st}}$, and the following relationship of output SCNR naturally holds:

$$\lim_{l\to+\infty}\text{SCNR}_{\text{Kro}} \approx \frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \sigma^2\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-2}\mathbf{p}_t\cdot\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-2}\mathbf{a}_{\text{st}}^{t}}. \tag{47}$$

Leveraging the Cauchy–Schwarz inequality, we have

$$\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\mathbf{a}_{\text{st}}^{t}\cdot\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-2}\mathbf{a}_{\text{st}}^{t} \geq \hat{\eta}_{\text{st}}^2 \tag{48}$$

$$\mathbf{p}_t^{\text{H}}\mathbf{p}_t\cdot\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-2}\mathbf{p}_t \geq \hat{\eta}_p^2 \tag{49}$$

then we have

$$\begin{aligned}&\frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \sigma^2\mathbf{p}_t^{\text{H}}\hat{\mathbf{R}}_p^{-2}\mathbf{p}_t\cdot\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-2}\mathbf{a}_{\text{st}}^{t}}\\ &\leq \frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \frac{\sigma^2}{\left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\mathbf{a}_{\text{st}}^{t}\cdot\mathbf{p}_t^{\text{H}}\mathbf{p}_t}\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2} = \frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \frac{\sigma^2}{\left\|\mathbf{a}_{\text{pst}}^{t}\right\|_2^2}\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}.\end{aligned} \tag{50}$$

The equation holds if and only if $\mathbf{a}_{\text{st}}^{t}$ ($\mathbf{p}_t$) is the eigenvector of $\hat{\mathbf{R}}_{\text{st}}$ ($\hat{\mathbf{R}}_p$). Thereby, for low-speed target detection, considering the high clutter power in actual airborne scenarios, even if the polarization difference between the target and clutter is large, leveraging the properties of standard Capon beamformer, we have $\hat{\eta}_{\text{st}} = \left(\mathbf{a}_{\text{st}}^{t}\right)^{\text{H}}\hat{\mathbf{R}}_{\text{st}}^{-1}\mathbf{a}_{\text{st}}^{t} \to 0$ and $\hat{\eta}_p\hat{\eta}_{\text{st}} \to 0$. Ultimately, the output SCNR of Kronecker estimators for low-speed targets is

$$\lim_{\substack{l\to+\infty\\ \hat{\eta}_p\hat{\eta}_{\text{st}}\to 0}}\text{SCNR}_{\text{Kro}} \leq \lim_{\hat{\eta}_p\hat{\eta}_{\text{st}}\to 0}\frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \frac{\sigma^2}{\left\|\mathbf{a}_{\text{pst}}^{t}\right\|_2^2}\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2} \to 0. \tag{51}$$

Besides, we can also easily derive the maximum output SCNR of Kronecker estimators, that is, for high-speed target detection with a large polarization difference ($\hat{\eta}_p\hat{\eta}_{\text{st}} \to +\infty$), having

$$\begin{aligned}\max\text{SCNR}_{\text{Kro}} &= \lim_{\substack{l\to+\infty\\ \hat{\eta}_p\hat{\eta}_{\text{st}}\to+\infty}}\text{SCNR}_{\text{Kro}}\\ &\leq \lim_{\hat{\eta}_p\hat{\eta}_{\text{st}}\to+\infty}\frac{\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}{\hat{\eta}_p\hat{\eta}_{\text{st}} + \frac{\sigma^2}{\left\|\mathbf{a}_{\text{pst}}^{t}\right\|_2^2}\hat{\eta}_p^2\hat{\eta}_{\text{st}}^2}\\ &= \lim_{\hat{\eta}_p\hat{\eta}_{\text{st}}\to+\infty}\frac{1}{\frac{1}{\hat{\eta}_p\hat{\eta}_{\text{st}}} + \frac{\sigma^2}{\left\|\mathbf{a}_{\text{pst}}^{t}\right\|_2^2}}\\ &= \frac{\left\|\mathbf{a}_{\text{pst}}^{t}\right\|_2^2}{\sigma^2} = \text{SNR}\end{aligned} \tag{52}$$

$$\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}\left[\ln p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})\middle|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2\right]\propto -LK\ln|\mathbf{P}|-\sum_{l=1}^{L}\text{Tr}\left(\left((\text{diag}(\mathbf{b}))^{-1}\otimes\mathbf{P}^{-1}\right)\mathbb{E}_{\boldsymbol{\beta}_l}\left[\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}\middle|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2\right]\right) \tag{54}$$

where $\text{SNR}=(\|\mathbf{a}_{\text{pst}}^t\|_2^2/\sigma^2)$ represents the output signal-to-noise ratio of the polarimetric array. It is important to note that (52) also verifies the rationality of the above derivation.

This completes the proof.

Appendix B
Derivation of (25)

Recall that $p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})\sim\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma})$, we have

$$\sum_{l=1}^{L}\ln p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})\propto -L\ln\left(\left(\prod_{k=1}^{K}b_k\right)^{N_p}|\mathbf{P}|^K\right)-\sum_{l=1}^{L}\text{Tr}\left(\left((\text{diag}(\mathbf{b}))^{-1}\otimes\mathbf{P}^{-1}\right)\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}\right). \tag{53}$$

Then, we can get (54), as shown at the top of the page. Among (54), we know that

$$\mathbb{E}_{\boldsymbol{\beta}_l}[\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2]=\int\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}\mathcal{CN}(\boldsymbol{\mu}_l,\boldsymbol{\Xi})d\boldsymbol{\beta}_l. \tag{55}$$

It is easy to attain that $\boldsymbol{\beta}_l-\boldsymbol{\mu}_l\sim\mathcal{CN}(\mathbf{0},\boldsymbol{\Xi})$. Then, (55) yields

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\beta}_l}[\boldsymbol{\beta}_l\boldsymbol{\beta}_l^{\text{H}}|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2]\\&=\mathbb{E}_{\boldsymbol{\beta}_l}\left[\left(\boldsymbol{\mu}_l+(\boldsymbol{\beta}_l-\boldsymbol{\mu}_l)\right)\left(\boldsymbol{\mu}_l+(\boldsymbol{\beta}_l-\boldsymbol{\mu}_l)\right)^{\text{H}}\middle|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2\right]\\&=\boldsymbol{\mu}_l\boldsymbol{\mu}_l^{\text{H}}+\mathbb{E}_{\boldsymbol{\beta}_l}\left[(\boldsymbol{\beta}_l-\boldsymbol{\mu}_l)(\boldsymbol{\beta}_l-\boldsymbol{\mu}_l)^{\text{H}}\middle|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2\right]\\&=\boldsymbol{\mu}_l\boldsymbol{\mu}_l^{\text{H}}+\boldsymbol{\Xi}.\end{aligned} \tag{56}$$

At this point, substituting (56) into (54), we get

$$\begin{aligned}&\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}[\ln p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2]\\&\propto -LK\ln|\mathbf{P}|\\&\quad-\sum_{l=1}^{L}\text{Tr}\left(\left((\text{diag}(\mathbf{b}))^{-1}\otimes\mathbf{P}^{-1}\right)(\boldsymbol{\mu}_l\boldsymbol{\mu}_l^{\text{H}}+\boldsymbol{\Xi})\right).\end{aligned} \tag{57}$$

By maximizing $\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}[\ln p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2]$, that is, let

$$\begin{aligned}&\frac{\partial\sum_{l=1}^{L}\mathbb{E}_{\boldsymbol{\beta}_l}[\ln p(\boldsymbol{\beta}_l;\mathbf{b},\mathbf{P})|\mathbf{x}_{\text{in},l};\mathbf{b},\mathbf{P},\sigma^2]}{\partial\mathbf{P}}\\&\triangleq -LK\mathbf{P}^{-1}+\sum_{l=1}^{L}\sum_{k=1}^{K}b_k^{-1}\mathbf{P}^{-1}\left(\boldsymbol{\mu}_l^k(\boldsymbol{\mu}_l^k)^{\text{H}}+\boldsymbol{\Xi}_k\right)\mathbf{P}^{-1}\\&=0\end{aligned} \tag{58}$$

where

$$\boldsymbol{\mu}_l^k=\boldsymbol{\mu}_{l,[N_p(k-1)+1:N_pk]} \tag{59}$$

$$\boldsymbol{\Xi}_k=\boldsymbol{\Xi}_{[N_p(k-1)+1:N_pk,N_p(k-1)+1:N_pk]}. \tag{60}$$

Finally, we can attain the updating rule of $\mathbf{P}$, i.e.,

$$\mathbf{P}=\frac{1}{LK}\sum_{l=1}^{L}\sum_{k=1}^{K}b_k^{-1}\left(\boldsymbol{\mu}_l^k(\boldsymbol{\mu}_l^k)^{\text{H}}+\boldsymbol{\Xi}_k\right). \tag{61}$$

This completes the derivation.

Appendix C
Discussion of Approximation (28)

Recall that

$$\begin{aligned}\mathbf{T}&=\mathbf{D}_{\text{in}}\boldsymbol{\Sigma}\mathbf{D}_{\text{in}}^{\text{H}}+\sigma^2\mathbf{I}_{N_D}\\&=\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}\right)\otimes\mathbf{P}+\sigma^2\mathbf{I}_{N_D}.\end{aligned} \tag{62}$$

If the CNR is high, the following approximate relationship is naturally established:

$$\mathbf{T}\approx\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}\right)\otimes\mathbf{P}. \tag{63}$$

At this point, we have

$$\begin{aligned}\mathbf{T}^{-1}&=\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}\right)^{-1}\otimes\mathbf{P}^{-1}\\&\approx\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}+\sigma^2\mathbf{I}_{MN}\right)^{-1}\otimes\mathbf{P}^{-1}.\end{aligned} \tag{64}$$

In addition, for any matrices $\mathbf{A}\in\mathbb{C}^{M\times M}$ and $\mathbf{I}_N$, we find that

$$\mathbf{A}\otimes\mathbf{I}_N+\mathbf{I}_{MN}=(\mathbf{A}+\mathbf{I}_M)\otimes\mathbf{I}_N. \tag{65}$$

Thereby, if the polarization correlation coefficient $|\rho_c|\to 0$ and power ratio $\chi_c=1$, that is $\mathbf{P}\to\mathbf{I}_{N_p}$. The following approximate relationship is also established:

$$\begin{aligned}\mathbf{T}&=\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}\right)\otimes\mathbf{P}+\sigma^2\mathbf{I}_{N_D}\\&\overset{\mathbf{P}\to\mathbf{I}_{N_p}}{\approx}\left(\mathbf{D}_{\text{st}}\text{diag}(\mathbf{b})\mathbf{D}_{\text{st}}^{\text{H}}+\sigma^2\mathbf{I}_{MN}\right)\otimes\mathbf{P}\end{aligned} \tag{66}$$

and the resulting $\mathbf{T}^{-1}$ is equal to that in (63).

For PSTAP applications, although $|\rho_c|\to 0$ is not expected, high CNR is common for practical airborne radar systems. Thereby, approximation (28) is quite reasonable.

This completes the discussion.

References

[1] Y. Li, J. Chen, and J. Zhu, "A new ground accelerating target imaging method for airborne CSSAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[2] J. Li et al., "Joint clutter suppression and moving target indication in 2-D azimuth rotated time domain for single-channel bistatic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5202516.

[3] J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Lab., Lexington, MA, USA, Tech. Rep. 1015, Dec. 1994.

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

[5] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA, USA: Artech House, 2014.

[6] D. Giuli, "Polarization diversity in radars," *Proc. IEEE*, vol. 74, no. 2, pp. 245–269, Feb. 1986.

[7] A. de Maio, G. Alfano, and E. Conte, "Polarization diversity detection in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 1, pp. 114–131, Jan. 2004.

[8] Y. Wang, W. Xia, Z. He, H. Li, and A. P. Petropulu, "Polarimetric detection in compound Gaussian clutter with Kronecker structured covariance matrix," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4562–4576, Sep. 2017.

[9] Q. Liu, X. Zhang, Y. Liu, K. Huo, W. Jiang, and X. Li, "Multi-polarization fusion few-shot HRRP target recognition based on meta-learning framework," *IEEE Sensors J.*, vol. 21, no. 16, pp. 18085–18100, Aug. 2021.

[10] S. Han, P. Addabbo, F. Biondi, C. Clemente, D. Orlando, and G. Ricci, "Innovative solutions based on the EM-algorithm for covariance structure detection and classification in polarimetric SAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 1, pp. 209–227, Feb. 2023.

[11] J. Wang, Y. Wang, Z. Wang, Z. He, and B. Xiong, "Polarization-space-time domain adaptive detection for the heterogeneous array," *Signal Process.*, vol. 210, Sep. 2023, Art. no. 109094.

[12] A. Aubry, V. Carotenuto, A. De Maio, and F. Fioranelli, "Testing stationarity and statistical independence of multistatic/polarimetric sea-clutter with application to NetRAD data," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5103415.

[13] G. A. Showman, W. L. Melvin, and M. Belenkii, "Performance evaluation of two polarimetric STAP architectures," in *Proc. IEEE Radar Conf.*, May 2003, pp. 59–65.

[14] D. Wu, Z. Xu, L. Zhang, Z. Xiong, and S. Xiao, "Performance analysis of polarization-space-time three-domain joint processing for clutter suppression in airborne radar," *Prog. Electromagn. Res.*, vol. 129, pp. 579–601, 2012.

[15] L. Xie, Z. He, J. Tong, J. Li, and H. Li, "Transmitter polarization optimization for space-time adaptive processing with diversely polarized antenna array," *Signal Process.*, vol. 169, Apr. 2020, Art. no. 107401.

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

[17] L. M. Novak, M. C. Burl, and W. W. Irving, "Optimal polarimetric processing for enhanced target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 234–244, Jan. 1993.

[18] 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.

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

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

[21] 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.

[22] 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.

[23] 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. Electron. Syst.*, vol. 50, no. 1, pp. 254–264, Jan. 2014.

[24] L. Xie, Z. He, J. Tong, and W. Zhang, "A recursive angle-Doppler channel selection method for reduced-dimension space-time adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 3985–4000, Oct. 2020.

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

[26] N. Song, R. C. de Lamare, M. Haardt, and M. Wolf, "Adaptive widely linear reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4003–4016, Aug. 2012.

[27] K. Zhao, Z. Liu, S. Shi, Y. Huang, and Y. Xu, "Random nyström approach to clutter subspace estimation for polarimetric space-time adaptive processing," *Electron.*, vol. 9, no. 9, p. 1488, 2020.

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

[29] 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.

[30] K. Greenewald, E. Zelnio, and A. H. Hero, "Robust SAR STAP via Kronecker decomposition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2612–2625, Dec. 2016.

[31] 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.

[32] A. Breloy, G. Ginolhac, Y. Gao, and F. Pascal, "MIMO filters based on robust rank-constrained Kronecker covariance matrix estimation," *Signal Process.*, vol. 187, Oct. 2021, Art. no. 108116.

[33] L. Xie, Z. He, J. Tong, T. Liu, J. Li, and J. Xi, "Regularized covariance estimation for polarization radar detection in compound Gaussian sea clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5109416.

[34] T. Wang, D. Xu, C. Hao, P. Addabbo, and D. Orlando, "Clutter edges detection algorithms for structured clutter covariance matrices," *IEEE Signal Process. Lett.*, vol. 29, pp. 642–646, 2022.

[35] P. Addabbo, S. Han, D. Orlando, and G. Ricci, "Learning strategies for radar clutter classification," *IEEE Trans. Signal Process.*, vol. 69, pp. 1070–1082, 2021.

[36] V. Carotenuto, A. De Maio, D. Orlando, and L. Pallotta, "Adaptive radar detection using two sets of training data," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1791–1801, Apr. 2018.

[37] L. Yan, C. Hao, P. Addabbo, D. Orlando, and A. Farina, "An improved adaptive radar detector based on two sets of training data," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[38] L. Yan, P. Addabbo, C. Hao, D. Orlando, and A. Farina, "New ECCM techniques against noiselike and/or coherent interferers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1172–1188, Apr. 2020.

[39] S. Yan, F. Lotfi, S. Chen, C. Hao, and D. Orlando, "Innovative two-stage radar detection architectures in adverse scenarios using two training data sets," *IEEE Signal Process. Lett.*, vol. 28, pp. 1165–1169, 2021.

[40] 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.

[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] 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, Apr. 2023.

[43] I. Ahmed, A. Khalil, I. Ahmed, and J. Frnda, "Sparse signal representation, sampling, and recovery in compressive sensing frameworks," *IEEE Access*, vol. 10, pp. 85002–85018, 2022.

[44] P. He, S. He, Z. Yang, and P. Huang, "An off-grid STAP algorithm based on local mesh splitting with bistatic radar system," *IEEE Signal Process. Lett.*, vol. 27, pp. 1355–1359, 2020.

[45] 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.

[46] C. Liu, T. Wang, S. Zhang, and B. Ren, "A clutter suppression algorithm via weighted ℓ_2-norm penalty for airborne radar," *IEEE Signal Process. Lett.*, vol. 29, pp. 1522–1525, 2022.

[47] Z. Yang, X. Li, H. Wang, and W. Jiang, "Adaptive clutter suppression based on iterative adaptive approach for airborne radar," *Signal Process.*, vol. 93, no. 12, pp. 3567–3577, Dec. 2013.

[48] 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, Jul. 2017.

[49] 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, Jan. 2017.

[50] 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.

[51] 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.

[52] 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, Nov. 2024, Art. no. 109585.

[53] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[54] D.-P. Xia, L. Zhang, T. Wu, and X.-D. Meng, "A clutter-suppression method for airborne bistatic polarization radar based on polarization-space-time adaptive processing," *Multidimensional Syst. Signal Process.*, vol. 33, no. 3, pp. 899–916, Sep. 2022.

[55] D. J. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, May 1992.

[56] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.
[57] S. Sen, "Low-rank matrix decomposition and spatio-temporal sparse recovery for STAP radar," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1510–1523, Dec. 2015.
[58] J. Xu, S. Zhu, and G. Liao, "Range ambiguous clutter suppression for airborne FDA-STAP radar," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1620–1631, Dec. 2015.
[59] J. Zhang, "Airborne polarization array radar space-time processing technology research," M.S. thesis, Chengdu School Inf. Commun. Eng., Univ. Electron. Sci. Technol. China, Chengdu, China, 2023.
[60] W. Yifeng, W. Tong, W. Jianxin, and D. Jia, "Robust training samples selection algorithm based on spectral similarity for space–time adaptive processing in heterogeneous interference environments," *IET Radar, Sonar Navigat.*, vol. 9, no. 7, pp. 778–782, Aug. 2015.
[61] Z. Jiang, X. Wang, G. Li, X.-P. Zhang, and Y. He, "Space-time adaptive processing by employing structure-aware two-level block sparsity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6386–6397, 2021.
[62] Q. Guo, L. Liu, M. Kaliuzhnyi, L. Chernogor, Y. Wang, and L. Qi, "Robust STAP of dictionary local adaptive filling and learning for nonstationary clutter suppression," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 2, pp. 1284–1298, Apr. 2024.

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 (NUC), Taiyuan, China, in 2021. He is currently pursuing 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 (UESTC), Chengdu, China.

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

Jiaheng Wang was born in Hebei, China, in 1997. He received the B.S. degree in electronic engineering from the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, where he is currently pursuing the Ph.D. degree in information and communication engineering.

His current research interests include space-time adaptive processing, adaptive target detection, and array signal processing.

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 at 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 awarded the Young Scientists Award for Excellence in Scientific Research by the International Union of Radio Science (URSI) in 2020. He received the Prize for Excellent Doctor Degree Dissertation of Chinese Institute of Electronics in 2020.

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 adaptive signal processing, cognitive radar, and multi-input multi-output (MIMO) radar.

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), Chengdu, China, in 1984, 1988, and 2000, respectively.

He is currently a Professor of 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 beam forming, the theory of multiple-input multiple-output (MIMO) communication, MIMO radar, and interference cancellation.