Knowledge-aided multi-dictionary block sparsity-aware STAP for airborne polarimetric conformal array radar

Yalong Wang, Jiaheng Wang, Xuejing Zhang *, Jun Li, Zishu He

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

ARTICLE INFO

ABSTRACT

Airborne polarimetric conformal array radar has gained increased attention recently. However, the conformal array configuration and polarization factors may lead to non-stationary clutter, which significantly degrades the performance of space–time adaptive processing (STAP), particularly in short-range clutter environments. In this paper, we introduce a knowledge-aided multi-dictionary block sparse Bayesian learning (KA-MDBSBL) algorithm to improve the clutter suppression performance. Using prior knowledge, the proposed algorithm designs multi-dictionary matrices for each training sample, rather than a single dictionary matrix for the cell under test (CUT). We take advantage of the identical clutter profile under each dictionary matrix. In the multi-dictionary case, we enforce shared sparsity in the clutter profiles. Additionally, we utilize the inherent block structure of the dictionary matrix to jointly recover clutter and noise power through a fast convergence learning framework. Subsequently, the clutter plus noise covariance matrix is reconstructed using precisely estimated clutter and noise power, along with the dictionary matrix corresponding to the CUT. Numerical simulations are included to demonstrate the effectiveness and superiority of the proposed algorithm.

1. Introduction

With the development of miniaturization of unmanned early warning platforms, the conformal array has attracted increased interest due to numerous advantages over uniform linear arrays (ULAs) and planar arrays. These advantages include aerodynamic shaping that aligns with the platform, larger array aperture, reduced payload and radar cross section (RCS), and others [1,2]. Unlike ULAs and planar arrays, where each element is aligned in a uniform manner, the elements of a conformal array are placed on the surface of the platform. Therefore, the antenna patterns of the conformal array are no longer consistent in the global coordinate system. This characteristic allows the conformal array to sense polarization information, thereby enhancing target detection performance. The above conclusion has been verified through theoretical analysis and simulations, see the recent work in [3].

Space–time adaptive processing (STAP) is a powerful technique for the moving target indicator (MTI) on motion platforms [4]. When STAP technology is applied in conformal arrays, there are some challenges that need to be addressed. The use of special antenna arrangements introduces non-linearity in array geometry, leading to non-stationary clutter distribution [5–8]. Additionally, the clutter polarization properties in different range cells exhibit non-uniformity due to the influence of antenna patterns and varying terrain scattering characteristics [9,

10]. As a consequence, the training samples obtained from the range cells adjacent to the cell under test (CUT) do not strictly adhere to the independently and identically distributed (IID) condition. This is particularly evident for short-range clutter, where available training samples are scarce.

To address the shortage of training samples, the reduced-dimension (RD) STAP utilizes a well-designed transformation matrix to convert the full-dimension data into the local domain data [11,12]. The reduced-rank (RR) STAP, on the other hand, projects the echo signal onto a low-dimensional subspace [13,14]. It is generally considered that RD-STAP reduces the training sample number to twice the local system degrees of freedom, while RR-STAP reduces it to twice the clutter rank. However, the design of auxiliary channels and the accurate determination of clutter rank for conformal arrays remain open challenges. To address the range-dependent training sample issue, numerous clutter compensation methods have been proposed. These include angle-Doppler compensation (ADC) [15], adaptive angle-Doppler compensation (A2DC) [16], and so on. These methods aim to align the peaks of the clutter ridge in the training samples as closely as possible to that of the CUT. Note that these methods are only effective in highly directional antenna scenarios, and offer only partial clutter compensation. Additionally, the direct data domain (D3) approach [17] exclusively utilizes the data from CUT

without requiring training samples, making it an appealing option for conformal arrays. However, this compromise leads to aperture loss, resulting in performance degradation.

Recently, sparse recovery (SR) technologies have been applied to STAP by exploiting the inherent sparsity of clutter and using only a small number of training samples to reconstruct the clutter plus noise covariance matrix (CNCM). These technologies primarily fall into three categories: greedy algorithms [18], convex/nonconvex optimization methods [19–21], and Bayesian learning techniques [22–24]. Among these approaches, sparse Bayesian learning (SBL) captures significant attention due to its self-regulating user parameters and robustness against correlation. Existing SBL-STAP methods typically rely on training samples to recover the CNCM based on a shared dictionary matrix. However, the shared dictionary matrix is not suitable for conformal array STAP due to the non-linear array geometry and different antenna patterns. In contrast, the multi-dictionary SBL approach uses multiple underlying dictionaries to process samples. This method has been studied for image processing [25], wideband direction-of-arrival estimation [26], etc, indicating its potential applicability in conformal array STAP.

In this paper, a knowledge-aid multi-dictionary block sparse Bayesian learning (KA-MDBSBL) method is presented to improve the performance of polarimetric conformal array SR-STAP. Firstly, the signal model and work mechanism are introduced in detail for an airborne polarimetric conformal array radar. Subsequently, we exploit prior knowledge, which includes airborne radar system parameters and conformal array configuration, to devise multi-dictionary matrices for training samples and the CUT. We further illustrate the multi-dictionary problem and unique dictionary block structure. Finally, a fast convergence iterative framework is derived to estimate the clutter and noise power, resulting in a precise CNCM. Numerical simulations are carried out to demonstrate the superiority of the KA-MDBSBL across various non-stationary clutter backgrounds.

The remainder of this paper is organized as follows. Section 2 introduces the signal model, work mechanism, and STAP principle for the airborne polarimetric conformal array radar. Section 3 discusses the conformal array SR-STAP problem. Section 4 analyzes the details of the proposed conformal array SR-STAP method. Section 5 provides the simulation results to verify the effectiveness of the proposed method. Section 6 concludes this paper.

Notations: In this paper, we use lightface for scalar a, lower case boldface for vector $\mathbf{a}$, and upper case for matrix $\mathbf{A}$. $\mathbf{A}^{-1}$, $\mathbf{A}_{:,q}$ denote the inverse and the qth column of the matrix $\mathbf{A}$, respectively. diag($\mathbf{a}$) is the diagonalization of the vector $\mathbf{a}$, and blkdiag($\cdot$) forms a block diagonal matrix. Double line body $\mathbb{C}$ means the complex numbers domain. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\otimes$, $\odot$, and $\mathbb{E}\{\cdot\}$ denote the conjugate, transpose, conjugate transpose, Kronecker product, Hadamard product, and statistical expectation operators, respectively. $\mathbf{I}_N$ is the N-dimensional identity matrix. $\mathcal{R}\{\cdot\}$, $|\cdot|$, and $\mathrm{Tr}(\cdot)$ are the range space, determinant, and trace of a matrix, respectively. $\|\cdot\|_2$, $\|\cdot\|_F$, and $\|\cdot\|_{2,0}$ are the l_2 norm, Frobenius norm, and $l_{2,0}$ mixed norm, respectively. $\propto$ means that the quantities on its both sides are directly proportional.

2. Signal model and problem formulation

2.1. Transmitted signal model

For the sake of illustration, we consider a semicircular conformal array, as shown in Fig. 1. It is important to note that the proposed model is still applicable to other conformal arrays. The conformal array platform flies along the x-axis with a constant velocity v and crab angle φ_b at the height H. Here we assume that between every two adjacent antennas, one dipole is installed parallel to the tangent direction of the semicircle, and another dipole is installed perpendicular to this direction. The arc length distance between adjacent antennas is $\lambda/2$, where λ stands for the radar wavelength. Suppose that the number of
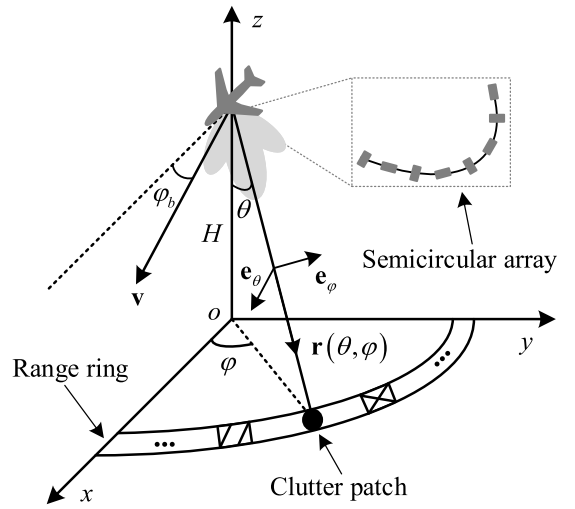

Fig. 1. Geometry model of airborne polarimetric conformal array radar.

elements is N, and we denote the position of the nth element in global coordinate system as $\mathbf{d}_n = [x_n, y_n, z_n]^T$, $n = 1, \ldots, N$. The unit vector describing the signal propagation direction is given by

$$\mathbf{r}(\theta, \varphi) = [\sin\theta\cos\varphi, \sin\theta\sin\varphi, -\cos\theta]^T \quad (1)$$

where θ and φ are the elevation angle and azimuth angle, respectively. Note that they are defined in the global coordinate, as illustrated in Fig. 1.

In contrast to the traditional ULA, each antenna pattern of the conformal array is unique due to the varied orientations of the antenna installation. Specifically, we can express the antenna pattern $\mathbf{g}_n(\theta, \varphi)$ as

$$\mathbf{g}_n(\theta, \varphi) = g_{\tilde{\theta}}(\tilde{\theta}_n, \tilde{\varphi}_n)\mathbf{e}_{\tilde{\theta}} + g_{\tilde{\varphi}}(\tilde{\theta}_n, \tilde{\varphi}_n)\mathbf{e}_{\tilde{\varphi}}, n = 1, 2, \ldots, N \quad (2)$$

where $(\tilde{\theta}_n, \tilde{\varphi}_n)$ denotes the corresponding direction of (θ, φ) in the local spherical coordinate system of the nth element (as shown in Fig. 2(a)), $g_{\tilde{\theta}}(\tilde{\theta}_n, \tilde{\varphi}_n)$ and $g_{\tilde{\varphi}}(\tilde{\theta}_n, \tilde{\varphi}_n)$ represent two polarized components under base vectors $\mathbf{e}_{\tilde{\theta}}$ and $\mathbf{e}_{\tilde{\varphi}}$, respectively. Since $g_{\tilde{\theta}}(\tilde{\theta}_n, \tilde{\varphi}_n)$ and $g_{\tilde{\varphi}}(\tilde{\theta}_n, \tilde{\varphi}_n)$ are usually defined in local coordinate system,[1] coordinate transformation is necessary to obtain the nth pattern $\mathbf{g}_n(\theta, \varphi)$ for the given (θ, φ). A concise coordinate transformation process is shown in Fig. 2(b). For more details, one can refer to [28].

Assuming that the antenna pattern $\mathbf{g}_n(\theta, \varphi)$ $(n = 1, 2, \ldots, N)$ is known, we further define $\mathbf{G}(\theta, \varphi)$ as

$$\mathbf{G}(\theta, \varphi) = [\mathbf{g}_1(\theta, \varphi), \mathbf{g}_2(\theta, \varphi), \ldots, \mathbf{g}_N(\theta, \varphi)]^T \in \mathbb{C}^{N \times 2} \quad (3)$$

Also, it is not difficult to obtain the following spatial steering vector

$$\mathbf{a}_s(\theta, \varphi) = [e^{-j2\pi f_{s,1}}, e^{-j2\pi f_{s,2}}, \ldots, e^{-j2\pi f_{s,N}}]^T \in \mathbb{C}^N \quad (4)$$

where $f_{s,n} = \mathbf{d}_n^T \mathbf{r}(\theta, \varphi)/\lambda, n = 1, 2, \ldots, N$. The manifold matrix $\mathbf{A}(\theta, \varphi)$ of the polarimetric conformal array can be denoted as

$$\mathbf{A}(\theta, \varphi) = \mathrm{diag}(\mathbf{a}_s(\theta, \varphi))\mathbf{G}(\theta, \varphi) = [\mathbf{a}_{s, \mathbf{e}_\theta}(\theta, \varphi), \mathbf{a}_{s, \mathbf{e}_\varphi}(\theta, \varphi)] \in \mathbb{C}^{N \times 2} \quad (5)$$

With the above notations and performing transmit beamforming using weight vector $\mathbf{w}_t$, we can express the synthesized polarization signal in direction (θ, φ) as [29]

$$\mathbf{p}(\theta, \varphi) = [\mathbf{w}_t^H \mathbf{A}(\theta, \varphi)]^T = [p_{\mathbf{e}_\theta}(\theta, \varphi), p_{\mathbf{e}_\varphi}(\theta, \varphi)]^T \in \mathbb{C}^2 \quad (6)$$

---

[1] One can refer to [27] to obtain the detailed patterns of different types of antennas.
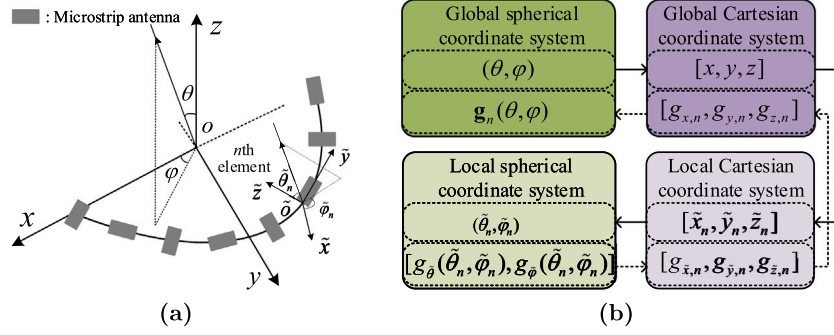
Fig. 2. (a) The semicircular array. (b) Transformation relationship between different coordinate systems.

where

$$p_{\mathbf{e}_\theta}(\theta,\varphi) = \mathbf{w}_t^H \mathbf{a}_{s,\mathbf{e}_\theta}(\theta,\varphi), \quad p_{\mathbf{e}_\varphi}(\theta,\varphi) = \mathbf{w}_t^H \mathbf{a}_{s,\mathbf{e}_\varphi}(\theta,\varphi) \tag{7}$$

According to (6), one can see that the radiation beampattern is related to $\mathbf{w}_t$, and any state of polarization can be synthesized by optimizing $\mathbf{w}_t$ [29]. Thus, the polarimetric conformal array owns multi-polarization characteristics, which can provide more information about the target and clutter.

2.2. Received signal model

We suppose that M pulses at a constant pulse repetition frequency (PRF) of f_r are transmitted during a coherent processing interval (CPI). The temporal (Doppler) steering vector is given by

$$\mathbf{a}_d(f_d^t) = [1, e^{-j2\pi f_d^t}, \dots, e^{-j2\pi(M-1)f_d^t}]^T \in \mathbb{C}^M \tag{8}$$

where $f_d^t = 2[v \sin\theta_t \cos(\varphi_t - \varphi_b) + v_t]/(\lambda f_r)$ denotes the normalized Doppler frequency, θ_t, φ_t and v_t are elevation angle, azimuth angle and radical velocity of the target, respectively. Then the polarization-space steering vector [30] of the target can be expressed as

$$\mathbf{a}_{ps}(\theta_t,\varphi_t,\mathbf{S}_t) = \mathbf{A}(\theta_t,\varphi_t)\mathbf{S}_t\mathbf{p}(\theta_t,\varphi_t) \in \mathbb{C}^N \tag{9}$$

where $\mathbf{S}_t$ is the scattering matrix and provides the polarization transforming characteristics of the target, its expression is given by

$$\mathbf{S}_t = \begin{bmatrix} s_{hh,t} & s_{hv,t} \\ s_{vh,t} & s_{vv,t} \end{bmatrix} \tag{10}$$

Assuming that the velocity, RCS, and polarization characteristics of the target are fixed in a CPI, one can express the noise-free polarization-space–time received signal of the target as

$$\mathbf{x}_t = \alpha_t \mathbf{a}_{ps}(\theta_t,\varphi_t,\mathbf{S}_t) \otimes \mathbf{a}_d(f_d^t) = \alpha_t \mathbf{a}_{psd} \in \mathbb{C}^{MN} \tag{11}$$

where α_t is the target complex reflection coefficient, $\mathbf{a}_{psd} = \mathbf{a}_{ps}(\theta_t,\varphi_t,\mathbf{S}_t) \otimes \mathbf{a}_d(f_d^t)$ denotes the target polarization-space–time steering vector.

Following the classical integral clutter model in [31], we assume that the interested area contains L_{all} range cells, and the successive azimuth angle φ is uniformly separated into N_c discrete angles, resulting in point-like clutter patch, as shown in Fig. 1. Without considering the range ambiguous clutter, the clutter echo of the lth range cell is expressed as the sum of N_c clutter patches, i.e.,

$$\mathbf{x}_c^l = \sum_{i=1}^{N_c} \alpha_c^{l,i} \mathbf{a}_{ps}(\theta_l,\varphi_i,\mathbf{S}_c^{l,i}) \otimes \mathbf{a}_d(f_{d,c}^{l,i}) \in \mathbb{C}^{NM} \tag{12}$$

where $\alpha_c^{l,i}$ denotes the complex reflection coefficient of the ith clutter patch for the lth range cell, $f_{d,c}^{l,i} = 2\mathbf{v}^T \mathbf{r}(\theta_l,\varphi_i)/(\lambda f_r)$ is the normalized Doppler frequency, $\mathbf{v} = [v \cos\varphi_b, v \sin\varphi_b, 0]^T$ represents the platform velocity vector, $\mathbf{S}_c^{l,i}$ represents the scattering matrix of the ith clutter patch in the lth range cell, i.e.,

$$\mathbf{S}_c^{l,i} = \begin{bmatrix} s_{hh,c}^{l,i} & s_{hv,c}^{l,i} \\ s_{vh,c}^{l,i} & s_{vv,c}^{l,i} \end{bmatrix}, \quad i = 1,\dots,N_c, \quad l = 1,\dots,L_{all} \tag{13}$$

By the reciprocity principle [9], we have $s_{hv,c}^{l,i} = s_{vh,c}^{l,i}$. Then $\mathbf{S}_c^{l,i}$ can be determined by the following three-element vector

$$\mathbf{s}_c^{l,i} = [s_{hh,c}^{l,i}, s_{hv,c}^{l,i}, s_{vv,c}^{l,i}]^T \tag{14}$$

In this paper, we assume that $\mathbf{s}_c^{l,i}$'s are IID and follow complex Gaussian distribution with zero-mean and covariance matrix $\mathbf{R}_p$ [32], i.e., $\mathbf{s}_c^{l,i} \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_p)$.

2.3. Problem formulation

For a given range CUT, the radar binary hypothesis testing problem can be formulated as

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{x}_c + \mathbf{n} \\ \mathcal{H}_1 : \mathbf{x} = \mathbf{x}_t + \mathbf{x}_c + \mathbf{n} \end{cases} \tag{15}$$

where we have omitted the range cell index l for the clutter echo. In (15), hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ correspond to target absence and presence, respectively. $\mathbf{n}$ represents the zero-mean Gaussian white noise with covariance matrix $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{MN}$, i.e., $\mathbf{n} \sim \mathcal{CN}(0,\sigma_n^2\mathbf{I}_{MN})$, and σ_n^2 is the noise power.

Assume that clutter, target signal, and noise are independent of each other, and the clutter from each patch is uncorrelated with each other. Then the target-free clutter plus noise covariance matrix can be written as

$$\begin{aligned} \mathbf{R} &= \mathbb{E}\{(\mathbf{x}_c + \mathbf{n})(\mathbf{x}_c + \mathbf{n})^H\} = \mathbf{R}_c + \mathbf{R}_n \\ &= \sum_{i=1}^{N_c} \mathbb{E}\{|\alpha_c^i|^2\} \mathbf{a}_{ps}(\theta,\varphi_i,\mathbf{S}_c^i)\mathbf{a}_{ps}^H(\theta,\varphi_i,\mathbf{S}_c^i) \otimes \mathbf{a}_d(f_{d,c}^i)\mathbf{a}_d^H(f_{d,c}^i) + \sigma_n^2\mathbf{I}_{MN} \end{aligned} \tag{16}$$

where $\mathbf{R}_c$ is the clutter covariance matrix. In practice, $\mathbf{R}$ is estimated using L training samples $\{\mathbf{x}_l\}_{l=1}^L$, i.e.,

$$\bar{\mathbf{R}} = \frac{1}{L}\sum_{l=1}^L \mathbf{x}_l\mathbf{x}_l^H \tag{17}$$

Based on the minimum variance distortionless response (MVDR) criterion, the well-known STAP weight vector can be computed via

$$\hat{\mathbf{w}} = \frac{\bar{\mathbf{R}}^{-1}\mathbf{a}_{psd}}{\mathbf{a}_{psd}^H \bar{\mathbf{R}}^{-1}\mathbf{a}_{psd}} \tag{18}$$

The Reed–Mallett–Brennan (RMB) rule [33] demonstrates that at least $L \geq 2MN$ IID training samples maintain the average loss of less than 3 dB. However, for conformal arrays, the clutter environments are usually non-stationary and heterogeneous. In this paper, we are concerned with the covariance matrix estimation problem for conformal arrays.

3. SR-STAP problem for conformal arrays

By exploiting the sparsity of clutter in the angle-Doppler domain, sparse recovery (SR) techniques can be utilized to improve the estimation of the clutter covariance matrix. In contrast to the regular planar arrays or ULAs, conformal arrays have non-uniform configurations and the radiation pattern of each antenna is different. Recalling (12), for a given range CUT, one can determine the corresponding elevation angle θ_0 and uniformly discretize K azimuth angles $\{\varphi_k\}_{k=1}^K$ in the azimuthal dimension. Then, to apply the SR technique, we can construct the following overcomplete dictionary matrix $\mathbf{D}_0$ for the polarimetric conformal array

$$\mathbf{D}_0 = [\mathbf{B}(\theta_0, \varphi_1, f_d^{0,1}), \mathbf{B}(\theta_0, \varphi_1, f_d^{0,2}), \dots, \mathbf{B}(\theta_0, \varphi_K, f_d^{0,K})] \in \mathbb{C}^{MN \times 2K^2} \quad (19)$$

where

$$\mathbf{B}(\theta_0, \varphi_p, f_d^{0,q}) = \mathbf{A}(\theta_0, \varphi_p) \otimes \mathbf{a}_d(f_d^{0,q}) \in \mathbb{C}^{MN \times 2}, \quad p, q = 1, 2, \dots, K \quad (20)$$

In (20), $f_d^{0,q} = 2\mathbf{v}^T\mathbf{r}(\theta_0, \varphi_q)/(\lambda f_r)$, $\mathbf{A}(\theta_0, \varphi_p)$ can be determined from the spatial steering vector $\mathbf{a}_s(\theta_0, \varphi_k)$ and the antenna pattern $\mathbf{g}_n(\theta_0, \varphi_k)$, $k = 1, 2, \dots, K$, $n = 1, 2, \dots, N$, as shown in (5).

Based on the principle of SR-STAP, the received signal of the range CUT can be rewritten as

$$\mathbf{x} = \mathbf{D}_0\boldsymbol{\gamma} + \mathbf{n} \quad (21)$$

and the estimation of the sparse vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_{2K^2}]^T$ can be formulated as

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_0 \quad \text{s.t.} \ \|\mathbf{x} - \mathbf{D}_0\boldsymbol{\gamma}\|_2^2 \leq \xi \quad (22)$$

where ξ is a given threshold. Several approximate algorithms can be used to find the solution to the above problem (22); see [19,21] for reference. Once $\boldsymbol{\gamma}$ and σ_n^2 are estimated, if the target is absent, the CNCM of CUT can be constructed via

$$\hat{\mathbf{R}} = \mathbf{D}_0 \text{diag}(|\gamma_1|^2, |\gamma_2|^2, \dots, |\gamma_{2K^2}|^2)\mathbf{D}_0^H + \sigma_n^2\mathbf{I}_{MN} \quad (23)$$

And the STAP weight vector can be calculated using (18).

In practice, if L IID training samples $\{\mathbf{x}_l\}_{l=1}^L$ are available, the CNCM of the CUT can be estimated using multiple measurement vector (MMV) model. In this case, the received signal matrix can be written as

$$\mathbf{X} = \mathbf{D}_0\boldsymbol{\Psi} + \mathbf{N} \quad (24)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L] \in \mathbb{C}^{MN \times L}$ is the training sample matrix, $\boldsymbol{\Psi} = [\tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2, \dots, \tilde{\boldsymbol{\gamma}}_L] \in \mathbb{C}^{2K^2 \times L}$ denotes the sparse coefficient matrix, and $\mathbf{N}$ is the Gaussian white noise matrix. Similarly, utilizing the inherent sparsity in $\boldsymbol{\Psi}$, the problem of coefficient matrix estimation can be formulated as [20,23]

$$\min_{\boldsymbol{\Psi}} \|\boldsymbol{\Psi}\|_{2,0} \quad \text{s.t.} \ \|\mathbf{X} - \mathbf{D}_0\boldsymbol{\Psi}\|_F^2 \leq L\xi \quad (25)$$

And the CNCM estimation of the CUT is calculated via

$$\hat{\mathbf{R}} = \frac{1}{L}\sum_{l=1}^L \mathbf{D}_0 \text{diag}(\boldsymbol{\Psi}_{:,l} \odot \boldsymbol{\Psi}_{:,l}^*)\mathbf{D}_0^H + \sigma_n^2\mathbf{I}_{MN} \quad (26)$$

In the aforementioned SR-STAP technique, the range-dependent characteristics of the training samples are not considered. For conformal arrays, a single dictionary matrix cannot fully represent the properties of multiple training samples. In addition, the echo of each clutter patch owns two polarized components under base vectors $\mathbf{e}_\theta$ and $\mathbf{e}_\varphi$, thus each atom in the dictionary matrix $\mathbf{D}_0$ is no longer a space–time steering vector, but a submatrix $\mathbf{B}(\theta_0, \varphi_p, f_d^{0,q})$ as indicated in (20), and K^2 submatrices form the dictionary matrix $\mathbf{D}_0$. Therefore, the dictionary matrix shows a block structure but is not utilized in the above-mentioned technique.

4. KA-MDBSBL STAP

In this section, we use multiple dictionaries to represent training samples and consider the inherent block sparse structure for the clutter covariance matrix. On this basis, we propose a multi-dictionary block sparse Bayesian learning STAP algorithm for the conformal array. The proposed algorithm has a fast convergence speed.

4.1. Represent training samples using multiple dictionaries

Different from the single dictionary SR-STAP formulation as given in (24)–(26), we use multiple dictionaries to represent the training samples adjacent to CUT and express $\{\mathbf{x}_l\}_{l=1}^L$ as

$$\mathbf{x}_l = \mathbf{D}_l\boldsymbol{\gamma}_l + \mathbf{n}_l, l = 1, 2, \dots, L \quad (27)$$

where $\mathbf{D}_l$ and $\boldsymbol{\gamma}_l = [\gamma_{l,1}, \gamma_{l,2}, \dots, \gamma_{l,2K^2}]^T$ stand for the dictionary matrix and sparse coefficient vector of the lth training sample, respectively. The construction of $\mathbf{D}_l$ is similar to $\mathbf{D}_0$, except for the difference in the selection of elevation angle.

One of the advantages of the multi-dictionary model can be demonstrated by using clutter ridges. Recall Fig. 1, ignore the effect of earth curvature, the relationship between the normalized Doppler frequency $f_{d,c}^{l,i}$, crab angle φ_b, elevation angle θ_l or slant range R_l (θ_l/R_l), and azimuth angle φ_i of the l, ith clutter patch can be expressed as

$$\begin{aligned} f_{d,c}^{l,i} &= \frac{2\mathbf{v}^T\mathbf{r}(\theta_l, \varphi_i)}{\lambda f_r} \\ &= v\sin\theta_l\cos(\varphi_i - \varphi_b) = v\sqrt{1 - \frac{H^2}{R_l^2}}\cos(\varphi_i - \varphi_b) \end{aligned} \quad (28)$$

From (28), we can see that θ_l/R_l is the key point resulting in range-dependent clutter. In the single-dictionary representation scenario, it is obvious that $f_{d,c}^{l,i}$ is a function of θ_l/R_l, φ_i, and φ_b, i.e., $\mathcal{F}(\theta_l/R_l, \varphi_i, \varphi_b)$, the normalized Doppler frequency range of clutter increases as θ_l/R_l increases, thus the clutter ridge distribution changes with the range-dependent samples. On the contrary, in the multi-dictionary representation scenario, we take the elevation information $\theta_l/R_l, l = 1, 2, \dots, L$ of samples into account, and then $f_{d,c}^{l,i}$ evolves into a function of φ_i and φ_b, i.e., $\mathcal{F}_{\theta_l/R_l}(\varphi_i, \varphi_b)$, thus clutter ridges occupy the same position under different dictionaries when multiple dictionaries of the same dimension (the value of K in each dictionary is identical) are applied. It is worth noting that φ_b corresponds to different looking modes, which only affects the shape of clutter ridges. For clarity, we also depict the clutter ridge distribution of samples at different slant ranges R_l in Fig. 3, where both the side-looking mode ($\varphi_b = 90°$) and the forward-looking mode ($\varphi_b = 0°$) are considered.

It should be noted that, due to the range-variant $\mathbf{D}_l, l = 1, 2, \dots, L$ in the multi-dictionary representation scenario, $\boldsymbol{\gamma}_l, l = 1, 2, \dots, L$ exhibit varying sparsities compared to the CUT, and each $\boldsymbol{\gamma}_l$ does not share the same sparsity, i.e.,

$$\mathbb{E}\{\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^H\} \neq \mathbb{E}\{\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^H\}, \ i, j = 1, \dots, L, \ i \neq j \quad (29)$$

Consequently, directly using estimates of multiple sparse coefficient vectors $\boldsymbol{\gamma}_l$'s to reconstruct the CNCM of CUT will be inaccurate.

Since the traditional clutter compensation algorithms only provide partial compensation and do not consider changes in the antenna pattern, the range dependence of training samples is still serious. Herein, from the perspective of SR, to extract and utilize as much of the shared sparse components between training samples and CUT as possible, we constrain the estimated clutter profiles (i.e., $\boldsymbol{\gamma}_l$) to share a common sparsity with the CUT, thus mitigating the impact of non-stationary clutter. Specifically, all the sparse coefficient vectors $\boldsymbol{\gamma}_l$'s are constrained to follow circularly symmetric complex Gaussian distribution with zero-mean and a shared covariance $\boldsymbol{\Xi}_0$. Under this
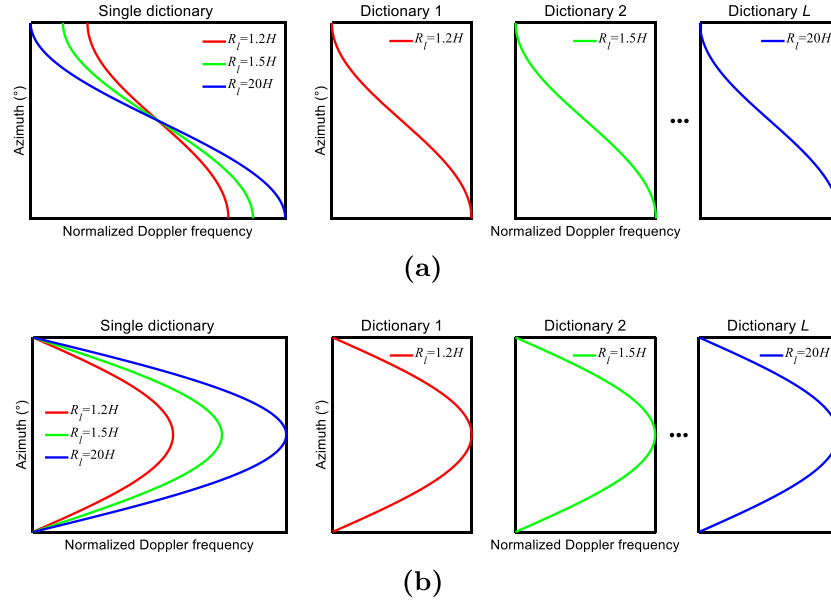
Fig. 3. Clutter ridge distribution in the case of single-dictionary and multi-dictionary representation. (a) Side-looking mode. (b) Forward-looking mode.
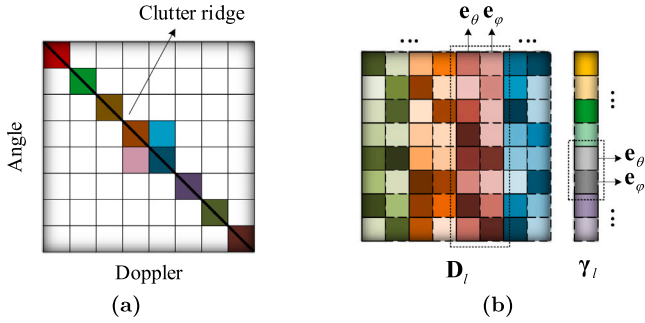

Fig. 4. (a) Clutter profile for a side-looking airborne radar. (b) Block structure of the dictionary matrix and the sparse coefficient vector.

constraint, the joint distribution of $\gamma_1, \gamma_2, \ldots, \gamma_L$ can thus be expressed as

$$p(\gamma_1, \gamma_2, \ldots, \gamma_L) = \prod_{l=1}^{L} p(\gamma_l) = \prod_{l=1}^{L} \mathcal{CN}(\gamma_l; \mathbf{0}, \Xi_0) \tag{30}$$

Subsequently, the CNCM of the CUT can be reconstructed via $\mathbf{D}_0$, Ξ_0, and σ_n^2, i.e.,

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{D}_0 \mathbb{E}\{\gamma_l \gamma_l^H\} \mathbf{D}_0^H + \sigma_n^2 \mathbf{I}_{MN} = \mathbf{D}_0 \Xi_0 \mathbf{D}_0^H + \sigma_n^2 \mathbf{I}_{MN} \tag{31}$$

This motivates us to use multi-dictionary matrices corresponding to training samples to jointly recover the clutter profile. In the following Section 4.3, we propose an algorithm that can quickly converge to estimate Ξ_0 and σ_n^2.

4.2. Clutter block sparse characteristic

Another important property of the airborne polarimetric conformal array is the clutter block sparsity, which is mainly reflected in two aspects. On one hand, the clutter profile behaves a special diagonal clustering structure along the clutter ridge, as shown in Fig. 4(a). Related work on improving the sparse recovery performance using the above-mentioned diagonal structure can be found in [34,35]. On the other hand, the clutter subspace $\mathbf{U}_{c,l}$ of the lth range cell is spanned

by $\mathbf{A}(\theta_l, \varphi_i) \otimes \mathbf{a}_d(f_{d,c}^{l,i})$ corresponding to the ith clutter patch ($i = 1, 2, \ldots, N_c$), i.e.,

$$\mathcal{R}\{\mathbf{U}_{c,l}\} = \mathcal{R}\{[\mathbf{A}(\theta_l, \varphi_1) \otimes \mathbf{a}_d(f_{d,c}^{l,1}), \ldots, \mathbf{A}(\theta_l, \varphi_{N_c}) \otimes \mathbf{a}_d(f_{d,c}^{l,N_c})]\} \tag{32}$$

Since $\mathbf{A}(\theta_l, \varphi_i) \in \mathbb{C}^{N \times 2}$ contains two polarized components on $\mathbf{e}_\theta$ and $\mathbf{e}_\varphi$, this leads to a block structure of the dictionary matrix and the sparse coefficient vector, as shown in Fig. 4(b).

To provide a unified description of single-dictionary and multi-dictionary scenarios, we re-express $\mathbf{D}_l$ and γ_l in (27) as

$$\mathbf{D}_l = [\mathbf{D}_{l,1}, \ldots, \mathbf{D}_{l,c}, \ldots, \mathbf{D}_{l,C}], \quad \gamma_l = [\gamma_{l,1}^T, \ldots, \gamma_{l,c}^T, \ldots, \gamma_{l,C}^T]^T \tag{33}$$

where $C = 2K^2/b$ is the total number of blocks, and b denotes the block size. In (33), $\gamma_{l,c} = [\gamma_{l,(c-1)b+1}, \ldots, \gamma_{l,cb}]^T \in \mathbb{C}^b$ and $\mathbf{D}_{l,c} = [(\mathbf{D}_l)_{:,(c-1)b+1}, \ldots, (\mathbf{D}_l)_{:,cb}] \in \mathbb{C}^{MN \times b}$ are the sub-vector of γ_l and the sub-dictionary matrix of $\mathbf{D}_l$, respectively. It is not difficult to find that the block SR problem degenerates into the conventional SR problem when $b = 1$. In the case of polarimetric conformal array, we set $b = 2$.

4.3. Multi-dictionary block sparse Bayesian learning STAP

According to the aforesaid multi-dictionary representation and block structure, we assume that $\gamma_{l,c}$'s are IID zero-mean Gaussian with covariance $\beta_c \Sigma_c$ for $\forall l \in \{1, \ldots, L\}$. Then, the probability density function (PDF) of $\gamma_{l,c}$ can be expressed as

$$p(\gamma_{l,c}; \beta_c, \Sigma_c) = \frac{1}{(\pi\beta_c)^b |\Sigma_c|} \exp\left(-\frac{\gamma_{l,c}^H \Sigma_c^{-1} \gamma_{l,c}}{\beta_c}\right) \tag{34}$$
$$l = 1, 2, \ldots, L, \quad c = 1, 2, \ldots, C$$

where β_c is the hyperparameter that controls the block sparsity of γ_l, $\Sigma_c \succ \mathbf{0}$ indicates the inner property of $\gamma_{l,c}$, $l = 1, \ldots, L$. Since the blocks are independent, the prior PDF of γ_l can be expressed as

$$p(\gamma_l; \Xi_0) = \prod_{c=1}^{C} p(\gamma_{l,c}; \beta_c, \Sigma_c) \tag{35}$$

where

$$\Xi_0 = \text{blkdiag}(\beta_1 \Sigma_1, \ldots, \beta_C \Sigma_C) \tag{36}$$

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_C]^T \tag{37}$$

Based on the Bayesian rule and using the fact that $\mathbf{n}_l \sim C\mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{MN})$, the posterior distribution over $\boldsymbol{\gamma}_l$ is given by

$$p(\boldsymbol{\gamma}_l | \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) = \frac{p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2) p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)}{\int p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2) p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0) d\boldsymbol{\gamma}_l} \quad (38)$$

where

$$p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2) = \frac{1}{(\pi \sigma_n^2)^{MN}} \exp\left(-\frac{\|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\gamma}_l\|_2^2}{\sigma_n^2}\right) \quad (39)$$

Since both $p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)$ and $p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2)$ follow Gaussian distributions, it can be derived that the probability density $p(\boldsymbol{\gamma}_l | \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0)$ is also Gaussian. Therefore, we denote the distribution $p(\boldsymbol{\gamma}_l | \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0)$ by

$$p(\boldsymbol{\gamma}_l | \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) \sim C\mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Xi}_l), \quad l = 1, 2, \dots, L \quad (40)$$

For a detailed derivation of $\boldsymbol{\mu}_l$ and $\boldsymbol{\Xi}_l$, one can see Appendix, and they take the form of

$$\boldsymbol{\mu}_l = \sigma_n^{-2} \boldsymbol{\Xi}_l \mathbf{D}_l^H \mathbf{x}_l = \boldsymbol{\Xi}_0 \mathbf{D}_l^H (\mathbf{D}_l \boldsymbol{\Xi}_0 \mathbf{D}_l^H + \sigma_n^2 \mathbf{I}_{MN})^{-1} \mathbf{x}_l \quad (41)$$

$$\boldsymbol{\Xi}_l = (\sigma_n^{-2} \mathbf{D}_l^H \mathbf{D}_l + \boldsymbol{\Xi}_0^{-1})^{-1} = \boldsymbol{\Xi}_0 - \boldsymbol{\Xi}_0 \mathbf{D}_l^H (\mathbf{D}_l \boldsymbol{\Xi}_0 \mathbf{D}_l^H + \sigma_n^2 \mathbf{I}_{MN})^{-1} \mathbf{D}_l \boldsymbol{\Xi}_0 \quad (42)$$

from which we can use $\boldsymbol{\mu}_l$ as an estimate for $\boldsymbol{\gamma}_l$. It is important to emphasize that $\boldsymbol{\Xi}_0$ is a common matrix and thus invariant in different formulas of $\boldsymbol{\mu}_l$ and $\boldsymbol{\Xi}_l$, $l = 1, 2, \dots, L$. We have enforced that different training samples share common sparsity under multi-dictionary representation, thus we can jointly use multiple dictionaries to recover $\boldsymbol{\Xi}_0$.

To do so, we let $\boldsymbol{\Gamma}_l = \mathbf{D}_l \boldsymbol{\Xi}_0 \mathbf{D}_l^H + \sigma_n^2 \mathbf{I}_{MN}$. Applying the matrix inversion lemma, one can readily obtain that

$$\boldsymbol{\Gamma}_l^{-1} = \sigma_n^{-2} \mathbf{I}_{MN} - \sigma_n^{-2} \mathbf{D}_l \boldsymbol{\Xi}_l \mathbf{D}_l^H \sigma_n^{-2} \quad (43)$$

Based on Appendix, the joint PDF of $\{\mathbf{x}_l\}_{l=1}^L$ is given by

$$\begin{aligned} p(\mathbf{X}; \sigma_n^2, \boldsymbol{\Xi}_0) &= \prod_{l=1}^L p(\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) = \prod_{l=1}^L \int p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2) p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0) d\boldsymbol{\gamma}_l \\ &= \prod_{l=1}^L \int C\mathcal{N}(\mathbf{x}_l; \mathbf{D}_l \boldsymbol{\gamma}_l, \sigma_n^2 \mathbf{I}_{MN}) C\mathcal{N}(\boldsymbol{\gamma}_l; \mathbf{0}, \boldsymbol{\Xi}_0) d\boldsymbol{\gamma}_l \\ &= \prod_{l=1}^L C\mathcal{N}(\mathbf{x}_l; \mathbf{0}, \boldsymbol{\Gamma}_l) \end{aligned} \quad (44)$$

On this basis, we can obtain the type-II maximum likelihood (ML) estimate of σ_n^2 and $\boldsymbol{\Xi}_0$ by solving the following problem

$$\{\hat{\sigma}_n^2, \hat{\boldsymbol{\Xi}}_0\} = \underset{\sigma_n^2, \boldsymbol{\Xi}_0}{\arg\min}\, C_0(\sigma_n^2, \boldsymbol{\Xi}_0) \quad (45)$$

with

$$\begin{aligned} C_0(\sigma_n^2, \boldsymbol{\Xi}_0) &= -\ln p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L; \sigma_n^2, \boldsymbol{\Xi}_0) \\ &= \sum_{l=1}^L \left(\ln|\boldsymbol{\Gamma}_l| + \mathbf{x}_l^H \boldsymbol{\Gamma}_l^{-1} \mathbf{x}_l\right) + LMN \ln \pi \end{aligned} \quad (46)$$

Substituting (43) into (46), using the fact that $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$, and omitting the constant term, we have

$$\begin{aligned} C_0(\sigma_n^2, \boldsymbol{\Xi}_0) &\propto \sum_{l=1}^L \left(\ln|\boldsymbol{\Gamma}_l| + \sigma_n^2 \|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\mu}_l\|_2^2 + \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\mu}_l\right) \\ &\triangleq C_1(\sigma_n^2, \boldsymbol{\Xi}_0) \end{aligned} \quad (47)$$

To solve the problem (45), we can apply the expectation maximization (EM) algorithm and take the derivatives of $C_1(\sigma_n^2, \boldsymbol{\Xi}_0)$ with respect to (w.r.t) $\boldsymbol{\Sigma}_c$, σ_n^2, and β_c. The updating rules for the calculations of $\boldsymbol{\Sigma}_c$, σ_n^2, and β_c can thus be obtained by setting the derivative as zero, respectively.

Considering the limited training samples, estimating $\boldsymbol{\Sigma}_c$ $\forall c$ will lead to overfitting [36,37], thus we constrain $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_C = \boldsymbol{\Sigma}$ to avoid overfitting. Note that the clutter-to-noise ratio (CNR) is relatively high

in practical clutter environments, this constraint does not affect the global sparsest solution [38]. According to (36), $\boldsymbol{\Xi}_0$ can be expressed as

$$\boldsymbol{\Xi}_0 = \text{diag}(\boldsymbol{\beta}) \otimes \boldsymbol{\Sigma} \quad (48)$$

From (44), we can notice that only $p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)$ is related to $\boldsymbol{\Sigma}$. Let

$$\partial \sum_{l=1}^L \mathbb{E}_{\boldsymbol{\gamma}_l} \left\{\ln p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0) \Big| \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0\right\}/\partial \boldsymbol{\Sigma} = 0 \quad (49)$$

the update of $\boldsymbol{\Sigma}$ can be easily obtained by

$$\boldsymbol{\Sigma}^{(t)} = \frac{1}{LC} \sum_{l=1}^L \sum_{c=1}^C \left(\boldsymbol{\Xi}_{l,c} + \boldsymbol{\mu}_{l,c} \boldsymbol{\mu}_{l,c}^H\right)/\beta_c^{(t-1)} \quad (50)$$

where $\boldsymbol{\mu}_{l,c} = [\mu_{l,(c-1)b+1}, \dots, \mu_{l,cb}]^T$, $\boldsymbol{\Xi}_{l,c} = \boldsymbol{\Xi}_l[(c-1)b+1 : cb, (c-1)b+1 : cb]$, and the superscript (t) represents the result of the (t)th iteration. For a lower computational complexity, one can ignore the slight impact of the intra-block correlation, and then constrain $\boldsymbol{\Sigma} = \mathbf{I}_b$.

Similarly, we can obtain the update rule of σ_n^2. Notice from (44) that only $p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2)$ depends on σ_n^2. Let

$$\partial \sum_{l=1}^L \mathbb{E}_{\boldsymbol{\gamma}_l} \left\{\ln p(\mathbf{x}_l | \boldsymbol{\gamma}_l; \sigma_n^2) \Big| \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0\right\}/\partial \sigma_n^2 = 0 \quad (51)$$

We can obtain the stationary point of σ_n^2 and the following update rule

$$(\sigma_n^2)^{(t)} = \frac{\sum_{l=1}^L \|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\mu}_l\|_2^2}{MNL - \sum_{l=1}^L \text{Tr}(\boldsymbol{\Xi}_0^{(t-1)} \mathbf{D}_l^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_l)} \quad (52)$$

For the update of β_c, the traditional EM algorithm has a slow learning rule, thus cannot be applied in real-time. To alleviate this issue, we refer to [22] and utilize the modified BSBL-BO algorithm to accelerate convergence for KA-MDBSBL STAP. Specifically, for each iteration step, the modified BSBL-BO algorithm minimizes the upper bound of the function $\sum_{l=1}^L \ln|\boldsymbol{\Gamma}_l|$ in $C_1(\sigma_n^2, \boldsymbol{\Xi}_0)$, using the fact that

$$\begin{aligned} \sum_{l=1}^L \ln|\boldsymbol{\Gamma}_l| &\leq \sum_{l=1}^L \ln|\boldsymbol{\Gamma}_l^{(t-1)}| \\ &+ \sum_{l=1}^L \sum_{c=1}^C \text{Tr}\left(\mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}\right)(\beta_c - \beta_c^{(t-1)}) \end{aligned} \quad (53)$$

Substituting (53) into (47), we obtain the following upper bound of $C_1(\sigma_n^2, \boldsymbol{\Xi}_0)$

$$\begin{aligned} C_2(\sigma_n^2, \boldsymbol{\Xi}_0) &\triangleq \sum_{l=1}^L \left(\ln|\boldsymbol{\Gamma}_l^{(t-1)}| + \sigma_n^2 \|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\mu}_l\|_2^2 + \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\mu}_l\right) \\ &+ \sum_{l=1}^L \sum_{c=1}^C \text{Tr}\left(\mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}\right)(\beta_c - \beta_c^{(t-1)}) \end{aligned} \quad (54)$$

After some calculations, we derive the derivative of $C_2(\sigma_n^2, \boldsymbol{\Xi}_0)$ w.r.t β_c as

$$\frac{\partial C_2(\sigma_n^2, \boldsymbol{\Xi}_0)}{\partial \beta_c} = \sum_{l=1}^L \left(\text{Tr}\left((\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma} \mathbf{D}_{l,c}^H\right) - \frac{\boldsymbol{\mu}_{l,c}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{l,c}}{\beta_c^2}\right) \quad (55)$$

Let $\partial C_2(\sigma_n^2, \boldsymbol{\Xi}_0)/\partial \beta_c = 0$, the update of β_c can thus be expressed as

$$\beta_c^{(t)} = \sqrt{\frac{\sum_{l=1}^L \boldsymbol{\mu}_{l,c}^H (\boldsymbol{\Sigma}^{(t-1)})^{-1} \boldsymbol{\mu}_{l,c}}{\sum_{l=1}^L \text{Tr}\left((\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}^{(t-1)} \mathbf{D}_{l,c}^H\right)}} \quad (56)$$

However, this update rule of β_c cannot intuitively reflect the change from $\beta_c^{(t-1)}$ to $\beta_c^{(t)}$, and the update speed is fixed. Immediately, substituting (41) into (56) yields

$$\beta_c^{(t)} = \beta_c^{(t-1)} \sqrt{\frac{\sum_{l=1}^L \mathbf{x}_l^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}^{(t-1)} \mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{x}_l}{\sum_{l=1}^L \text{Tr}\left(\mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}^{(t-1)}\right)}} \quad (57)$$

It is obvious that a self-regulating coefficient term controls the update process from $\beta_c^{(t-1)}$ to $\beta_c^{(t)}$, which is similar to the learning rate in machine learning. Since the self-regulating coefficient term in (57) always appears as the root form, the update speed is still fixed. To further control the speed of convergence, we introduce a parameter h to modify the learning rule for β_c, so that the self-regulating coefficient term is not limited to a specific form, i.e.,

$$\beta_c^{(t)} = \beta_c^{(t-1)} \left(\frac{\sum_{l=1}^{L} \mathbf{x}_l^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}^{(t-1)} \mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{x}_l}{\sum_{l=1}^{L} \mathrm{Tr} \left(\mathbf{D}_{l,c}^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{D}_{l,c} \boldsymbol{\Sigma}^{(t-1)} \right)} \right)^h \quad (58)$$

Repeating the above iteration steps, we can ultimately obtain the estimates of $\boldsymbol{\Xi}_0$ and σ_n^2. The above iteration process terminates until reaching the specified number of iterations (denoted as $t_{\max}$) or the following condition is satisfied

$$\|\boldsymbol{\Xi}_0^{(t)} - \boldsymbol{\Xi}_0^{(t-1)}\|_F^2 / \|\boldsymbol{\Xi}_0^{(t)}\|_F^2 \leq \tau \quad (59)$$

where τ is the preset convergence threshold.

Finally, the CNCM of the CUT can be reconstructed via (31) using the estimated $\boldsymbol{\Xi}_0$ and σ_n^2. The proposed KA-MDBSBL STAP algorithm is summarized in Algorithm 1.

Algorithm 1 KA-MDBSBL conformal array STAP

Input: $\{\mathbf{x}_l\}_{l=1}^{L}$, $\{\mathbf{D}_l\}_{l=1}^{L}$, h, $t_{\max}$ or τ
1: **Initialize:** $\beta_c = 1 \ \forall c$, $\sigma_n^2 = 10$, $\boldsymbol{\Sigma} = \mathbf{I}_b$, and $\boldsymbol{\Xi}_0 = \mathrm{diag}(\boldsymbol{\beta}) \otimes \mathbf{I}_b$
2: **Repeat**
3: Compute $\boldsymbol{\Gamma}_l^{(t-1)} = \mathbf{D}_l \boldsymbol{\Xi}_0^{(t-1)} \mathbf{D}_l^H + (\sigma_n^2)^{(t-1)} \mathbf{I}_{MN}$
4: Compute $\boldsymbol{\mu}_l^{(t-1)} = \boldsymbol{\Xi}_0^{(t-1)} \mathbf{D}_l^H (\boldsymbol{\Gamma}_l^{(t-1)})^{-1} \mathbf{x}_l$
5: Update $(\sigma_n^2)^{(t)}$ via (52)
6: Update $\beta_c^{(t)}$ via (58)
7: **Until** $t = t_{\max}$ or (59) is satisfied
8: Compute $\hat{\mathbf{R}}$ via (31)
Output: $\hat{\mathbf{R}}$

4.4. Computational complexity

In this subsection, we analyze the computational complexity of the proposed KA-MDBSBL STAP algorithm. In each iteration, we constrain $\boldsymbol{\Sigma} = \mathbf{I}_b$ by the above analysis, thus omitting the iterative process of $\boldsymbol{\Sigma}$, resulting in $\boldsymbol{\Xi}_0$ a diagonal matrix. Then, the computation complexity of the proposed algorithm is dominated by the updating rule for $\boldsymbol{\beta}$, which consists of two parts. The first part is computing $\boldsymbol{\Gamma}_l$ and $\boldsymbol{\Gamma}_l^{-1}$, with computational complexities $\mathcal{O}((KMN)^2)$ and $\mathcal{O}((MN)^3)$, respectively. The second part is the update in (58) with a computational complexity $\mathcal{O}(K^2 MN(MN + b + 1))$. Therefore, the overall computation complexity in each iteration is $\mathcal{O}(L(K^2 MN(MN + b + 1) + (KMN)^2 + (MN)^3))$. Compared to the classical fast converging SBL (FCSBL) STAP algorithm [24] (with a computational complexity $\mathcal{O}(K^2 MN(MN + L) + (KMN)^2 + (MN)^3)$), the proposed algorithm has an increased computational complexity.

5. Simulations

In this section, we present various simulations to assess the performance of the proposed method. The echo signal is generated via the model mentioned in Section 2. Specifically, each range cell contains 361 clutter patches, and the clutter polarization covariance matrix $\mathbf{R}_p$ is determined following the Ref. [9]. We assume that the target is located at the mainlobe axis (θ_0, φ_0) with scattering matrix $\mathbf{S}_t = \begin{bmatrix} 0.5j & -j \\ -j & 0.5 \end{bmatrix}$. Radar system parameters are listed in Table 1.

To represent the influence of clutter non-stationary, we consider two typical looking modes (side-looking with $\varphi_b = 90°$ and forward-looking with $\varphi_b = 0°$) at two different slant ranges R_t ($1.5H$ and $8H$),

Table 1
Radar system parameters.

Parameters	Values	Parameters	Values
Platform altitude	6 km	Platform velocity	150 m/s
Wavelength	0.3 m	Bandwidth	5 MHz
PRF	2000 Hz	Pulse number	8
Element number	11	CNR	40 dB

Table 2
The average computational time of SR-based algorithms.

Algorithm	Global runtime (s)	Single iteration runtime (s)
FCSBL	17.2324	0.8206
BSBL	32.7121	1.0552
KA-MDBSBL	76.2943	3.6331

respectively. For comparison, conventional sample matrix inversion (SMI) method and the ADC method [15] are considered. The number of training samples for these two methods is set to $L = 2MN$. Additionally, several SBL methods, such as the classical FCSBL algorithm [24] and block SBL (BSBL) algorithm [22] with block size $b = 2$, are also tested for comparison. For the proposed KA-MDBSBL method, we employ the prior knowledge of the array manifold $\mathbf{A}(\theta, \varphi)$ and the elevation angle $\theta_l, l = 1, 2, \ldots, L$ corresponding to range cells where training samples are located, uniformly discretize K azimuth angles $\{\varphi_k\}_{k=1}^{K}$ in the azimuthal dimension and construct multi-dictionary matrices $\mathbf{D}_l, l = 1, 2, \ldots, L$ similar to (19). Unless otherwise stated, the number of training samples and discrete azimuth angles for SR-based methods is chosen as $L = 8$ and $K = 6N$. The initialization parameters are set as follows: $\beta_c = 1, \forall c$, $\sigma_n^2 = 10$, $\boldsymbol{\Sigma} = \mathbf{I}_b$, $t_{\max} = 50$, $\tau = 10^{-3}$, and $h = 1$. The output signal-to-clutter-plus-noise ratio (SCNR) loss is adopted to quantify the performance of clutter cancellation, which is defined as follows

$$\mathrm{SCNR \ loss} = \frac{\sigma_n^2 \left| \hat{\mathbf{w}}^H \mathbf{a}_{psd} \right|^2}{\hat{\mathbf{w}}^H \mathbf{R} \hat{\mathbf{w}} \cdot \mathbf{a}_{psd}^H \mathbf{a}_{psd}} \quad (60)$$

Besides, the average SCNR loss is depicted under different training sample numbers. In this case, we take the average value of the SCNR loss over the entire normalized Doppler frequency range. All results are averages from 100 Monte-Carlo trials.

5.1. Computational efficiency of SR-based algorithms

In the first experiment, we present the computational time of SR-based methods, including average global runtime until convergence and average single iteration runtime, as shown in Table 2. For a fair comparison, the experiment results of SR-based methods are obtained by using MATLAB R2021b with the same frame. The hardware configuration is CPU AMD Ryzen 9 7845HX with Radeon Graphics 3 GHz and RAM 16 GB.

As we can see, FCSBL has the shortest computational time, while KA-MDBSBL has the longest computational time. However, according to the following simulation experiments, the excellent clutter suppression performance of KA-MDBSBL is the most suitable for the polarimetric conformal array among these algorithms. To improve the computational efficiency of the proposed method, we will consider parallel processing and extend KA-MDBSBL to the tensor-based framework in future work.

5.2. Performance comparison in convergence and clutter spectrum estimation

In this simulation, we first evaluate the convergence performance of SR-based algorithms. Specifically, the cost functions for KA-MDBSBL and FCSBL/BSBL are $\frac{1}{L} \sum_{l=1}^{L} \left(\ln |\boldsymbol{\Gamma}_l| + \mathrm{Tr}(\boldsymbol{\Gamma}_l^{-1} \mathbf{x}_l \mathbf{x}_l^H) \right)$ and $\ln |\boldsymbol{\Gamma}_0| +$
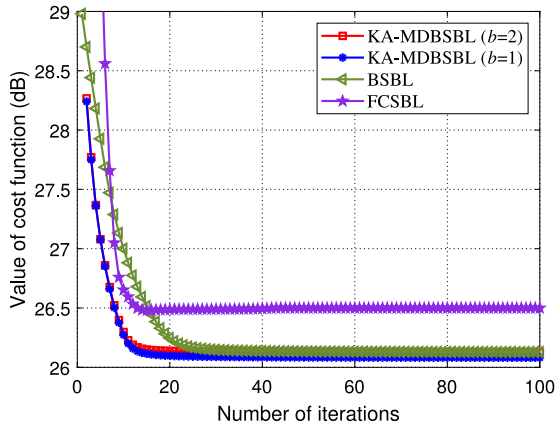
Fig. 5. Value of cost function versus the number of iterations.

$\mathrm{Tr}(\mathbf{\Gamma}_0^{-1}\bar{\mathbf{R}})$ [24], respectively. Fig. 5 shows curves of cost function value versus the iteration number, in the side-looking case with $R_t = 8H$. As we can see, all of the algorithms tested can converge within 30 iterations, and the KA-MDBSBL nearly has the fastest convergence speed. The FCSBL has almost the same convergence speed as KA-MDBSBL, but the final cost function value of FCSBL is higher than that of KA-MDBSBL and BSBL. Although BSBL has the slowest convergence speed, BSBL converges to almost the same value as KA-MDBSBL. The above simulation results confirm the effectiveness and fast convergence of the proposed algorithm.

Next, we compare the MVDR spectrum estimation accuracy of different algorithms to demonstrate clutter suppression performance for short-range clutter. Figs. 6 and 7 show the MVDR spectrums in side-looking mode and forward-looking mode, respectively. It is clear that the MVDR spectrums of SMI and ADC deviate significantly from the clairvoyant one, especially in forward-looking mode. On the contrary, the SR-based methods achieve a higher CNCM estimation precision, with KA-MDBSBL being closest to the clairvoyant one. Due to the influence of non-stationary clutter and the geometry of conformal arrays, the clutter spectrum estimation accuracy of FCSBL method is

unsatisfactory although it has a fast convergence speed. By considering the block structure of the dictionary matrix and sparse coefficient vector, the MVDR spectrum estimation accuracy of BSBL is better than that of FCSBL. The above simulation results demonstrate the superior performance of the proposed KA-MDBSBL algorithm in CNCM estimation for conformal arrays.

5.3. Performance comparison in output SCNR loss

To essentially assess the clutter suppression performance of the above algorithms, Figs. 8 and 9 exhibits the SCNR loss curves under different clutter non-stationary situations. Specifically, Fig. 8 provides the SCNR loss curves for various algorithms when $R_t = 1.5H$. Due to their dependence on the number of samples ($L = 2MN$), both ADC and SMI suffer from significant performance loss. Although BSBL and FCSBL reconstruct CNCM through fewer samples ($L = 8$), they still cannot achieve satisfactory clutter suppression performance due to the heterogeneity of training samples. The heterogeneity is particularly severe for the forward-looking mode.

Fig. 9 provides the SCNR loss curves for various algorithms when $R_t = 8H$, a case where the clutter non-stationary is slight. In this case, ADC and SMI have an approximately 3 dB average output SCNR loss compared to the optimal result (black curve). Although the clutter suppression performance of BSBL and FCSB has improved, there is still a gap between their performance and the optimal one due to the presence of non-stationary (weaker) clutter.

On the contrary, the proposed KA-MDBSBL algorithm designs multiple dictionaries based on prior knowledge to reconstruct the CNCM, which ensures that the clutter ridges overlap under the multi-dictionary representation. Meanwhile, the intrinsic sparsity of all training samples is controlled by the same second-order statistic to combat the non-stationary clutter. Consequently, the KA-MDBSBL algorithm achieves the clutter suppression performance closest to the optimal result. In addition, due to the consideration of the block structure of the dictionary matrix and sparse coefficient vector, the clutter suppression performance of KA-MDBSBL when $b = 2$ is better than that when $b = 1$, as shown in Figs. 8 and 9.

In order to further verify the impact of block size b on clutter suppression performance, Fig. 10 provides the SCNR loss curves of
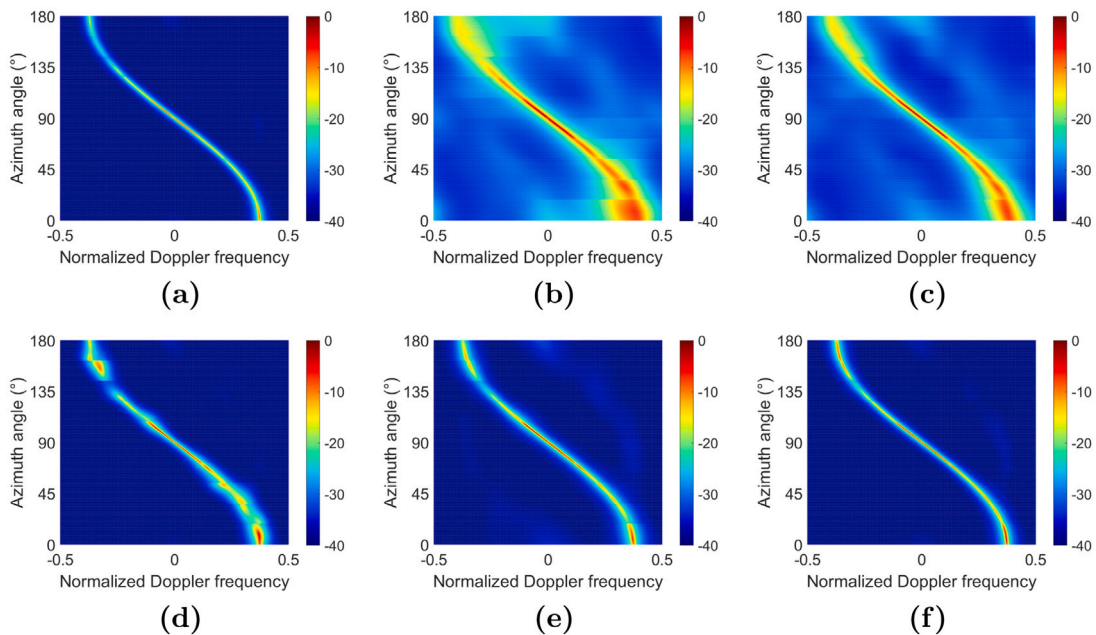

Fig. 6. Short-range clutter MVDR spectrums estimated by various methods in side-looking mode. (a) Clairvoyant MVDR spectrum. (b) SMI. (c) ADC. (d) FCSBL. (e) BSBL. (f) KA-MDBSBL with $b = 2$.
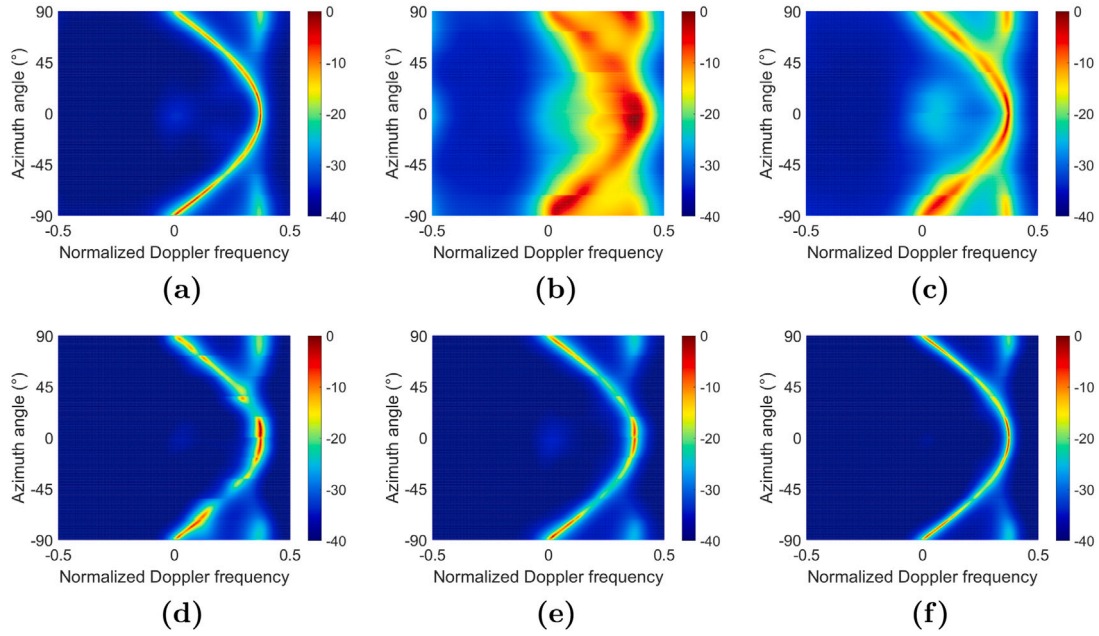
Fig. 7. Short-range clutter MVDR spectrums estimated by various methods in forward-looking mode. (a) Clairvoyant MVDR spectrum. (b) SMI. (c) ADC. (d) FCSBL. (e) BSBL. (f) KA-MDBSBL with $b = 2$.
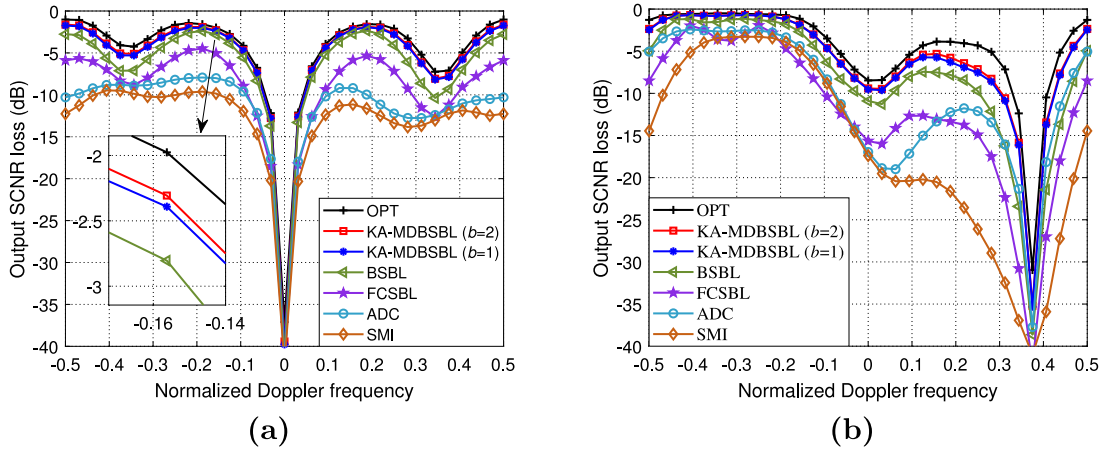

Fig. 8. The SCNR loss curves of various methods with $R_t = 1.5H$ under the ideal case. (a) Side-looking mode. (b) Forward-looking mode.
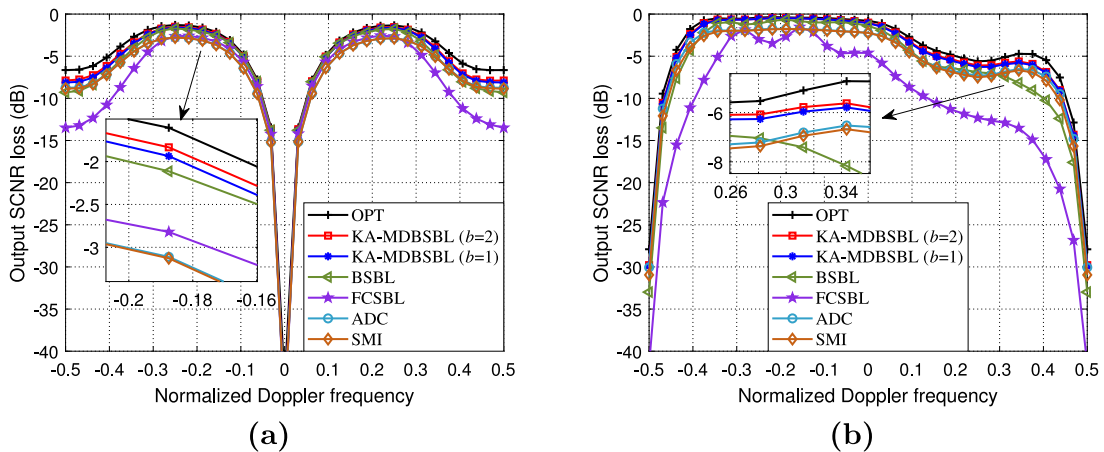

Fig. 9. The SCNR loss curves of various methods with $R_t = 8H$ under the ideal case. (a) Side-looking mode. (b) Forward-looking mode.
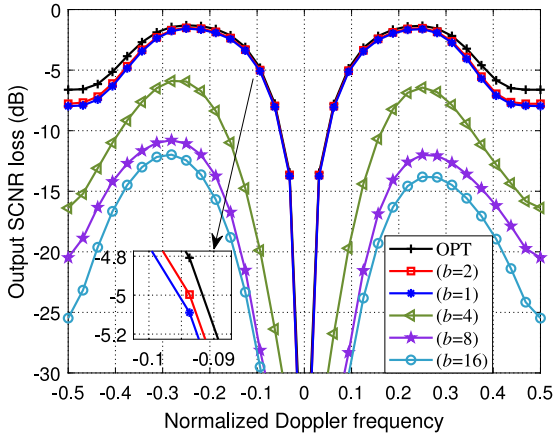
Fig. 10. The SCNR loss curves of KA-MDBSBL with different block sizes (side-looking mode, $R_t = 8H$).

KA-MDBSBL for different block sizes when $R_t = 8H$. It is evident that only when the block size matches the intrinsic structure of the dictionary matrix and sparse coefficient vector (i.e., $b = 2$) can we achieve satisfactory performance.

5.4. Average SCNR loss with different sample sizes

Now we evaluate the impact of the training sample number on the clutter suppression performance. Figs. 11 and 12 show the performance curves for $R_t = 1.5H$ under forward-looking mode and $R_t = 8H$ under side-looking mode, respectively. Specifically, Figs. 11(a) and 12(a) depict the curves of average SCNR loss, while Figs. 11(b) and 12(b) show the curves of average SCNR loss deviation (relative to the optimal value).

According to the results in Figs. 11 and 12, ADC and SMI converge quite slowly, while ADC owns a better performance as the number of training samples increases, especially in the case when the clutter non-stationary is severe. For the SBL-based methods, BSBL and FCSBL gradually converge as the number of training samples increases in the slight clutter non-stationary case. However, BSBL and FCSBL do not converge as the number of training samples increases in the severe non-stationary case. For the proposed KA-MDBSBL algorithm, it converges quickly in both cases. As the number of training samples continues to increase, the performance of KA-MDBSBL gradually approaches the optimal one. The above results validate the effectiveness of the proposed algorithm in suppressing non-stationary clutter.

5.5. Performance with non-ideal factors

In the last simulation, we investigated the impact of non-ideal factors on the proposed algorithm. We consider two common non-ideal factors, including angle-independent array gain-phase (GP) error [11, 12] and intrinsic clutter motion (ICM).

Specifically, the GP error is modeled as

$$\mathbf{e}_s = [(1 - \varepsilon_{a,1})e^{j\varepsilon_{b,1}}, \dots, (1 - \varepsilon_{a,N})e^{j\varepsilon_{b,N}}]^T \tag{61}$$

where $\varepsilon_{a,n}$ and $\varepsilon_{b,n}$, $n = 1, \dots, N$ follow a uniform distribution within $[0, \varepsilon_{a,\max}]$ and $[-\varepsilon_{b,\max}/2, \varepsilon_{b,\max}/2]$, respectively. Fig. 13(a) presents the performance in the forward-looking case with $R_t = 1.5H$ and $(\varepsilon_{a,\max}, \varepsilon_{b,\max}) = (3\%, \pi/36)$.

In the ICM case, we consider the Gaussian model proposed by Ward in [31], having

$$\mathbf{T} = \text{Toeplitz}(e_t(0), e_t(1), \dots, e_t(M - 1)) \tag{62}$$

$$e_t(m) = \exp\left\{-\frac{8\pi^2\sigma_v^2 T_r^2 m^2}{\lambda^2}\right\}, m = 0, 1, \dots, M - 1 \tag{63}$$

where σ_v^2 denotes the velocity standard deviation and $T_r = 1/f_r$ is the pulse repetition interval. Fig. 13(b) presents the performance in the forward-looking case with $R_t = 1.5H$ and $\sigma_v^2 = 1$.

Obviously, the mismatch between the dictionary and the data causes the performance deterioration for SBL-based algorithms. As expected, KA-MDBSBL presents the least deterioration and is closest to the optimal result. The above results illustrate that the KA-MDBSBL can achieve far better performance than other methods and is more suitable for conformal arrays.

6. Conclusions

In this paper, we have proposed a fast converging sparsity-aware STAP approach named KA-MDBSBL for the polarimetric conformal array. First, the signal model for airborne radar with a polarimetric conformal array has been introduced. Next, the prior knowledge of conformal array configuration and radar system parameters is utilized to design multi-dictionary matrices for training samples and the CUT, and carefully analyze the multi-dictionary problem. Meantime, the block structure of the dictionary matrix and sparse coefficient vector are also exploited to recover the clutter and noise power. Finally, the clutter plus noise covariance matrix is reconstructed from the precisely estimated clutter and noise power, as well as the dictionary matrix corresponding to the CUT. Simulation results demonstrate that the proposed algorithm can effectively improve the output performance under different non-stationary clutter environments.

CRediT authorship contribution statement

Yalong Wang: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Jiaheng Wang:** Writing – review & editing, Validation, Formal analysis. **Xuejing Zhang:** Writing – review & editing, Supervision, Project administration. **Jun Li:** Writing – review & editing. **Zishu He:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

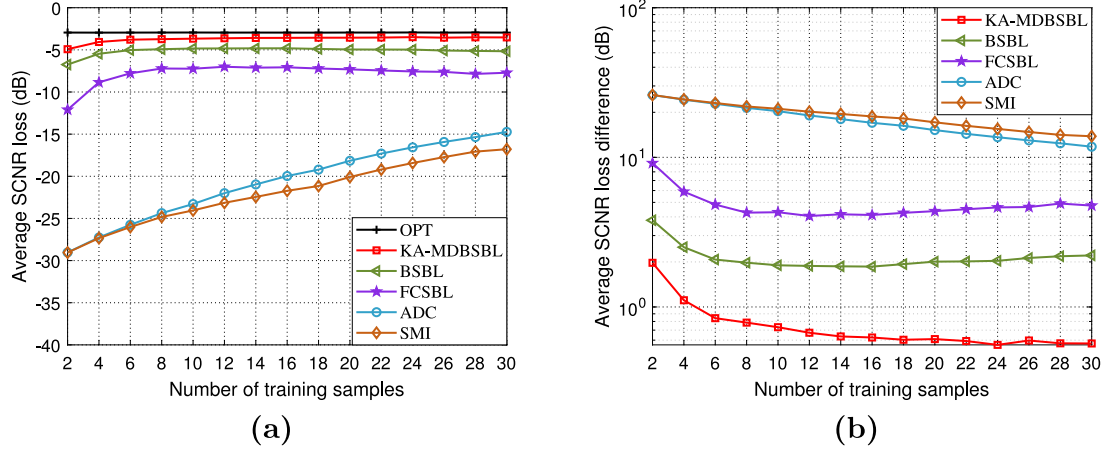The data that has been used is confidential.

Fig. 11. The average SCNR loss curves of various methods in forward-looking mode and $R_t = 1.5H$. (a) Average SCNR loss versus the number of training samples. (b) Average SCNR loss difference versus the number of training samples.
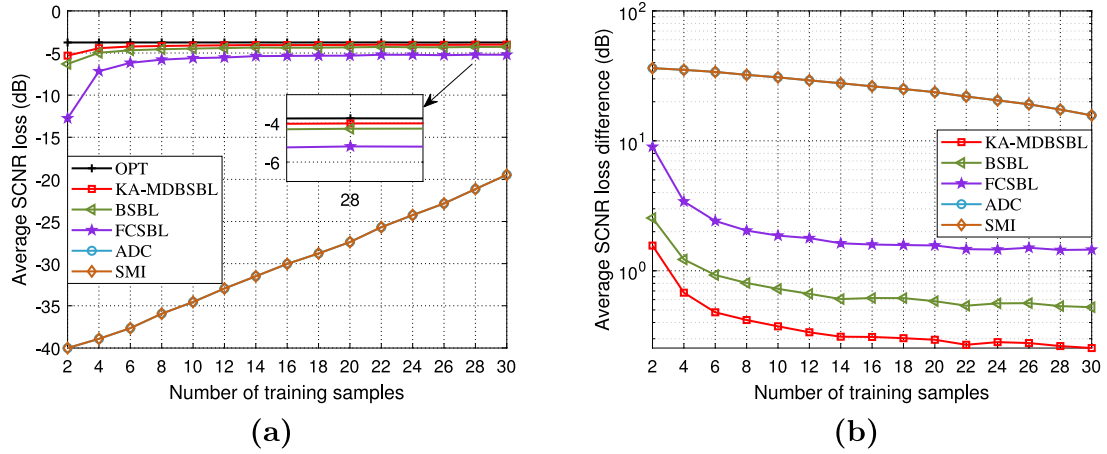
Fig. 12. The average SCNR loss curves of various methods in side-looking mode and $R_t = 8H$. (a) Average SCNR loss versus the number of training samples. (b) Average SCNR loss difference versus the number of training samples.
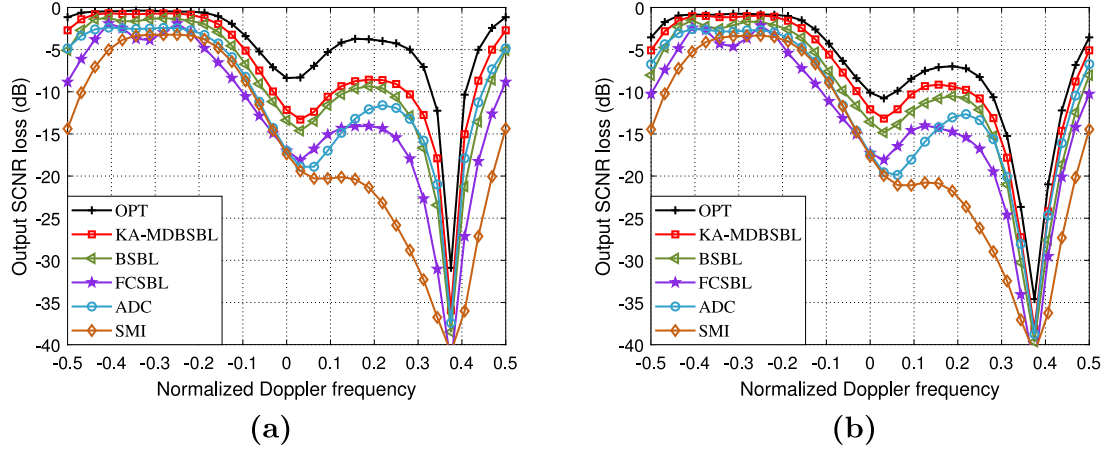
Fig. 13. The SCNR loss curves of various methods in forward-looking mode and $R_t = 1.5H$ under the non-ideal case. (a) Array GP error case. (b) ICM case.

Appendix. Derivation of the posterior distribution with multi-dictionary representation

Based on the Bayesian rule, having

$$p(\boldsymbol{\gamma}_l \mid \mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) = \frac{p(\mathbf{x}_l \mid \boldsymbol{\gamma}_l; \sigma_n^2)p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)}{\int p(\mathbf{x}_l \mid \boldsymbol{\gamma}_l; \sigma_n^2)p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)d\boldsymbol{\gamma}_l} \quad \text{(A.1)}$$

where

$$p(\mathbf{x}_l \mid \boldsymbol{\gamma}_l; \sigma_n^2) = \frac{1}{(\pi\sigma_n^2)^{MN}} \exp\left(-\frac{\|\mathbf{x}_l - \mathbf{D}_l\boldsymbol{\gamma}_l\|_2^2}{\sigma_n^2}\right) \quad \text{(A.2)}$$

and

$$p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0) = \frac{1}{\pi^{MN} |\boldsymbol{\Xi}_0|} \exp\left(-\boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\gamma}_l\right) \tag{A.3}$$

For notational simplicity, omitting the constant term, then the numerator term in (A.1) can be expressed as

$$p(\mathbf{x}_l|\boldsymbol{\gamma}_l; \sigma_n^2)p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0) = p(\mathbf{x}_l, \boldsymbol{\gamma}_l; \sigma_n^2, \boldsymbol{\Xi}_0)$$
$$\propto \exp\left(-\sigma_n^{-2} \|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\gamma}_l\|_2^2 - \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\gamma}_l\right) \tag{A.4}$$

The denominator term in (A.1) is well-known as the marginal likelihood, and takes the form of

$$p(\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) = \int p(\mathbf{x}_l|\boldsymbol{\gamma}_l; \sigma_n^2)p(\boldsymbol{\gamma}_l; \boldsymbol{\Xi}_0)d\boldsymbol{\gamma}_l$$
$$\propto \int \exp\left(-\sigma_n^{-2} \|\mathbf{x}_l - \mathbf{D}_l \boldsymbol{\gamma}_l\|_2^2 - \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\gamma}_l\right)d\boldsymbol{\gamma}_l$$
$$= \int \exp\left(-\left(\frac{\mathbf{x}_l^H \mathbf{x}_l - \boldsymbol{\gamma}_l^H \mathbf{D}_l^H \mathbf{x}_l - \mathbf{x}_l^H \mathbf{D}_l \boldsymbol{\gamma}_l + \boldsymbol{\gamma}_l^H \mathbf{D}_l^H \mathbf{D}_l \boldsymbol{\gamma}_l}{\sigma_n^2} + \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_0^{-1} \boldsymbol{\gamma}_l\right)\right)d\boldsymbol{\gamma}_l \tag{A.5}$$

Combining quadratic terms related to $\boldsymbol{\gamma}_l$, let $\boldsymbol{\Xi}_l = \left(\sigma_n^{-2}\mathbf{D}_l^H\mathbf{D}_l + \boldsymbol{\Xi}_0^{-1}\right)^{-1}$ and use the fact that $\mathbf{I}_{2K^2} = \boldsymbol{\Xi}_l\boldsymbol{\Xi}_l^{-1}$, (A.5) yields

$$p(\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0)$$
$$\propto \int \exp\left(-\left(\frac{\mathbf{x}_l^H \mathbf{x}_l - \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\Xi}_l \mathbf{D}_l^H \mathbf{x}_l - \mathbf{x}_l^H \mathbf{D}_l \boldsymbol{\Xi}_l \boldsymbol{\Xi}_l^{-1} \boldsymbol{\gamma}_l}{\sigma_n^2} + \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_l^{-1} \boldsymbol{\gamma}_l\right)\right)d\boldsymbol{\gamma}_l \tag{A.6}$$

Let $\boldsymbol{\mu}_l = \sigma_n^{-2}\boldsymbol{\Xi}_l\mathbf{D}_l^H\mathbf{x}_l$ and introduce item $\boldsymbol{\mu}_l^H\boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l$, (A.6) yields

$$p(\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0)$$
$$\propto \int \exp\left(-\left(\frac{\mathbf{x}_l^H \mathbf{x}_l}{\sigma_n^2} - \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l - \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\gamma}_l + \boldsymbol{\gamma}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\gamma}_l + \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l - \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l\right)\right)d\boldsymbol{\gamma}_l$$
$$= \int \exp\left(-\left(\frac{\mathbf{x}_l^H \mathbf{x}_l}{\sigma_n^2} - \boldsymbol{\mu}_l^H \boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l + (\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)^H \boldsymbol{\Xi}_l^{-1}(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)\right)\right)d\boldsymbol{\gamma}_l \tag{A.7}$$

Substituting $\boldsymbol{\mu}_l = \sigma_n^{-2}\boldsymbol{\Xi}_l\mathbf{D}_l^H\mathbf{x}_l$ into $\frac{\mathbf{x}_l^H \mathbf{x}_l}{\sigma_n^2} - \boldsymbol{\mu}_l^H\boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l$ and combining (43), we have

$$\frac{\mathbf{x}_l^H \mathbf{x}_l}{\sigma_n^2} - \boldsymbol{\mu}_l^H\boldsymbol{\Xi}_l^{-1}\boldsymbol{\mu}_l = \frac{\mathbf{x}_l^H \mathbf{x}_l}{\sigma_n^2} - \sigma_n^{-2}\mathbf{x}_l^H\mathbf{D}_l\boldsymbol{\Xi}_l\boldsymbol{\Xi}_l^{-1}\boldsymbol{\Xi}_l\mathbf{D}_l^H\mathbf{x}_l\sigma_n^{-2}$$
$$= \mathbf{x}_l^H\left(\sigma_n^{-2}\mathbf{I}_{MN} - \sigma_n^{-2}\mathbf{D}_l\boldsymbol{\Xi}_l\mathbf{D}_l^H\sigma_n^{-2}\right)\mathbf{x}_l$$
$$= \mathbf{x}_l^H\boldsymbol{\Gamma}_l^{-1}\mathbf{x}_l \tag{A.8}$$

Substituting (A.8) into (A.7) and using the fact that the integral of the probability density over its domain is equal to 1, (A.6) yields

$$p(\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) \propto \int \exp\left(-\left(\mathbf{x}_l^H\boldsymbol{\Gamma}_l^{-1}\mathbf{x}_l + (\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)^H \boldsymbol{\Xi}_l^{-1}(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)\right)\right)d\boldsymbol{\gamma}_l$$
$$= \exp\left(-\mathbf{x}_l^H\boldsymbol{\Gamma}_l^{-1}\mathbf{x}_l\right)\int \exp\left(-(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)^H \boldsymbol{\Xi}_l^{-1}(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)\right)d\boldsymbol{\gamma}_l$$
$$= \exp\left(-\left(\mathbf{x}_l^H\boldsymbol{\Gamma}_l^{-1}\mathbf{x}_l\right)\right)$$
$$\propto \mathcal{CN}\left(\mathbf{x}_l; \mathbf{0}, \boldsymbol{\Gamma}_l\right) \tag{A.9}$$

According to (A.9), (A.4) yields

$$p(\mathbf{x}_l, \boldsymbol{\gamma}_l; \sigma_n^2, \boldsymbol{\Xi}_0) \propto \exp\left(-\mathbf{x}_l^H\boldsymbol{\Gamma}_l^{-1}\mathbf{x}_l\right)\exp\left(-(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)^H \boldsymbol{\Xi}_l^{-1}(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)\right) \tag{A.10}$$

Substituting (A.9) and (A.10) into (A.1), we finally obtain

$$p(\boldsymbol{\gamma}_l|\mathbf{x}_l; \sigma_n^2, \boldsymbol{\Xi}_0) \propto \exp\left(-(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)^H \boldsymbol{\Xi}_l^{-1}(\boldsymbol{\gamma}_l - \boldsymbol{\mu}_l)\right)$$
$$\propto \mathcal{CN}\left(\boldsymbol{\mu}_l, \boldsymbol{\Xi}_l\right) \tag{A.11}$$

where $\boldsymbol{\mu}_l = \sigma_n^{-2}\boldsymbol{\Xi}_l\mathbf{D}_l^H\mathbf{x}_l$ and $\boldsymbol{\Xi}_l = \left(\sigma_n^{-2}\mathbf{D}_l^H\mathbf{D}_l + \boldsymbol{\Xi}_0^{-1}\right)^{-1}$. Furthermore, using the fact that $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$, $\boldsymbol{\mu}_l$ yields

$$\boldsymbol{\mu}_l = \sigma_n^{-2}\boldsymbol{\Xi}_l\mathbf{D}_l^H\mathbf{x}_l$$
$$= \sigma_n^{-2}\left(\sigma_n^{-2}\mathbf{D}_l^H\mathbf{D}_l + \boldsymbol{\Xi}_0^{-1}\right)^{-1}\mathbf{D}_l^H\mathbf{x}_l$$
$$= \left(\mathbf{D}_l^H\mathbf{D}_l + \sigma_n^2\boldsymbol{\Xi}_0^{-1}\right)^{-1}\mathbf{D}_l^H\mathbf{x}_l$$
$$= \left(\sigma_n^2\boldsymbol{\Xi}_0^{-1}\left(\mathbf{I}_{2K^2} + \sigma_n^{-2}\boldsymbol{\Xi}_0\mathbf{D}_l^H\mathbf{D}_l\right)\right)^{-1}\mathbf{D}_l^H\mathbf{x}_l$$
$$= \sigma_n^{-2}\left(\mathbf{I}_{2K^2} + \sigma_n^{-2}\boldsymbol{\Xi}_0\mathbf{D}_l^H\mathbf{D}_l\right)^{-1}\boldsymbol{\Xi}_0\mathbf{D}_l^H\mathbf{x}_l$$
$$= \sigma_n^{-2}\boldsymbol{\Xi}_0\mathbf{D}_l^H\left(\mathbf{I}_{MN} + \sigma_n^{-2}\mathbf{D}_l\boldsymbol{\Xi}_0\mathbf{D}_l^H\right)^{-1}\mathbf{x}_l$$
$$= \boldsymbol{\Xi}_0\mathbf{D}_l^H\left(\mathbf{D}_l\boldsymbol{\Xi}_0\mathbf{D}_l^H + \sigma_n^2\mathbf{I}_{MN}\right)^{-1}\mathbf{x}_l \tag{A.12}$$

Applying the matrix inversion lemma, $\boldsymbol{\Xi}_l$ yields

$$\boldsymbol{\Xi}_l = \left(\sigma_n^{-2}\mathbf{D}_l^H\mathbf{D}_l + \boldsymbol{\Xi}_0^{-1}\right)^{-1}$$
$$= \boldsymbol{\Xi}_0 - \boldsymbol{\Xi}_0\mathbf{D}_l^H(\mathbf{D}_l\boldsymbol{\Xi}_0\mathbf{D}_l^H + \sigma_n^2\mathbf{I}_{MN})^{-1}\mathbf{D}_l\boldsymbol{\Xi}_0 \tag{A.13}$$

References

[1] R.K. Hersey, Adaptive Detection and Estimation using a Conformal Array Antenna, Georgia Institute of Technology, 2004.

[2] R. Hersey, W. Melvin, J.H. McClellan, Clutter-limited detection performance of multi-channel conformal arrays, Signal Process. 84 (9) (2004) 1481–1500.

[3] J. Wang, Y. Wang, Z. Wang, Z. He, B. Xiong, Polarization-space–time domain adaptive detection for the heterogeneous array, Signal Process. 210 (2023) 109094.

[4] W.L. Melvin, A STAP overview, IEEE Aerosp. Electron. Syst. Mag. 19 (1) (2004) 19–35.

[5] Y. Sun, W.-q. Wang, Airborne flexible-array radar clutter characteristics and STAP performance analysis, IEEE Trans. Aerosp. Electr. Syst. (2023).

[6] F. Tao, T. Wang, J. Wu, Y. Su, A knowledge aided SPICE space time adaptive processing method for airborne radar with conformal array, Signal Process. 152 (2018) 54–62.

[7] K. Duan, W. Xie, Y. Wang, Nonstationary clutter suppression for airborne conformal array radar, Sci. China Inf. Sci. 54 (2011) 2170–2177.

[8] K. Sun, H. Meng, F.D. Lapierre, X. Wang, Registration-based compensation using sparse representation in conformal-array STAP, Signal Process. 91 (10) (2011) 2268–2276.

[9] L.M. Novak, M.C. Burl, W. Irving, Optimal polarimetric processing for enhanced target detection, IEEE Trans. Aerosp. Electr. Syst. 29 (1) (1993) 234–244.

[10] W.L. Melvin, Space–time adaptive radar performance in heterogeneous clutter, IEEE Trans. Aerosp. Electr. Syst. 36 (2) (2000) 621–633.

[11] W. Zhang, Z. He, J. Li, H. Liu, Y. Sun, A method for finding best channels in beam-space post-Doppler reduced-dimension STAP, IEEE Trans. Aerosp. Electr. Syst. 50 (1) (2014) 254–264.

[12] Y.-L. Wang, J.-W. Chen, Z. Bao, Y.-N. Peng, Robust space–time adaptive processing for airborne radar in nonhomogeneous clutter environments, IEEE Trans. Aerosp. Electr. Syst. 39 (1) (2003) 70–81.

[13] A. Haimovich, M. Berin, Eigenanalysis-based space–time adaptive radar: Performance analysis, IEEE Trans. Aerosp. Electr. Syst. 33 (4) (1997) 1170–1179.

[14] M.L. Honig, J.S. Goldstein, Adaptive reduced-rank interference suppression based on the multistage wiener filter, IEEE Trans. Commun. 50 (6) (2002) 986–994.

[15] B. Himed, Y. Zhang, A. Hajjari, STAP with angle-Doppler compensation for bistatic airborne radars, in: Proceedings of the 2002 IEEE Radar Conference (IEEE Cat. No. 02CH37322), IEEE, 2002, pp. 311–317.

[16] A.G. Jaffer, P.T. Ho, B. Himed, Adaptive compensation for conformal array STAP by configuration parameter estimation, in: 2006 IEEE Conference on Radar, IEEE, 2006, 6–pp.

[17] T.K. Sarkar, H. Wang, S. Park, R. Adve, J. Koh, K. Kim, Y. Zhang, M.C. Wicks, R.D. Brown, A deterministic least-squares approach to space–time adaptive processing (STAP), IEEE Trans. Antennas and Propagation 49 (1) (2001) 91–103.

[18] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.

[19] Z. Yang, R.C. de Lamare, X. Li, L_1-Regularized STAP algorithms with a generalized sidelobe canceler architecture for airborne radar, IEEE Trans. Signal Process. 60 (2) (2011) 674–686.

[20] C. Liu, T. Wang, S. Zhang, B. Ren, A clutter suppression algorithm via weighted ℓ_2−norm penalty for airborne radar, IEEE Signal Process. Lett. 29 (2022) 1522–1525.

[21] Z. Yang, X. Li, H. Wang, W. Jiang, Adaptive clutter suppression based on iterative adaptive approach for airborne radar, Signal Process. 93 (12) (2013) 3567–3577.

[22] Z. Zhang, B.D. Rao, Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation, IEEE Trans. Signal Process. 61 (8) (2013) 2009–2015.

[23] K. Duan, Z. Wang, W. Xie, H. Chen, Y. Wang, Sparsity-based STAP algorithm with multiple measurement vectors via sparse Bayesian learning strategy for airborne radar, IET Signal Process. 11 (5) (2017) 544–553.

[24] Z. Wang, W. Xie, K. Duan, Y. Wang, Clutter suppression algorithm based on fast converging sparse Bayesian learning for airborne radar, Signal Process. 130 (2017) 159–168.

[25] S. Ji, D. Dunson, L. Carin, Multitask compressive sensing, IEEE Trans. Signal Process. 57 (1) (2008) 92–106.

[26] S. Nannuru, K.L. Gemba, P. Gerstoft, W.S. Hodgkiss, C.F. Mecklenbräuker, Sparse Bayesian learning with multiple dictionaries, Signal Process. 159 (2019) 159–170.

[27] C.A. Balanis, Antenna Theory: Analysis and Design, John wiley & sons, 2016.

[28] T. Milligan, More applications of euler rotation angles, IEEE Antennas Propag. Mag. 41 (4) (1999) 78–83.

[29] B. Fuchs, J.J. Fuchs, Optimal polarization synthesis of arbitrary arrays with focused power pattern, IEEE Trans. Antennas and Propagation 59 (12) (2011) 4512–4519.

[30] X. Wang, S. Zhou, W. Zhao, X. Peng, H. Ma, Target polarization scattering matrix estimation with conformal MIMO radar, Signal Process. 210 (2023) 109054.

[31] J. Ward, Space–Time Adaptive Processing for Airborne Radar, Technical Report, Lincoln Lab. MIT, 1994.

[32] Y. Wang, W. Xia, Z. He, H. Li, A.P. Petropulu, Polarimetric detection in compound Gaussian clutter with kronecker structured covariance matrix, IEEE Trans. Signal Process. 65 (17) (2017) 4562–4576.

[33] I.S. Reed, J.D. Mallett, L.E. Brennan, Rapid convergence rate in adaptive arrays, IEEE Trans. Aerosp. Electr. Syst. (6) (1974) 853–863.

[34] Z. Jiang, X. Wang, G. Li, X.-P. Zhang, Y. He, Space–time adaptive processing by employing structure-aware two-level block sparsity, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 6386–6397.

[35] N. Cui, K. Xing, K. Duan, Z. Yu, Knowledge-aided block sparse Bayesian learning STAP for phased-array MIMO airborne radar, IET Radar Sonar Navig. 15 (12) (2021) 1628–1642.

[36] G.C. Cawley, N.L. Talbot, Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters, J. Mach. Learn. Res. 8 (4) (2007).

[37] I. Guyon, A. Saffari, G. Dror, G. Cawley, Model selection: beyond the Bayesian/frequentist divide, J. Mach. Learn. Res. 11 (1) (2010).

[38] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 912–926.