Article

An Improved Iterative Reweighted STAP Algorithm for Airborne Radar

Weichen Cui [ID], Tong Wang *, Degen Wang and Cheng Liu

National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China
* Correspondence: twang@mail.xidian.edu.cn

Abstract: In recent years, sparse recovery-based space-time adaptive processing (SR-STAP) technique has exhibited excellent performance with insufficient samples. Sparse Bayesian learning algorithms have received considerable attention for their remarkable and reliable performance. Its implementation in large-scale radar systems is however hindered by the overwhelming computational load and slow convergence speed. This paper aims to address these drawbacks by proposing an improved iterative reweighted sparse Bayesian learning algorithm based on expansion-compression variance-components (ExCoV-IIR-MSBL). Firstly, a modified Bayesian probabilistic model for SR-STAP is introduced. Exploiting the intrinsic sparsity prior of the clutter, we divide the space-time coefficients into two parts: the significant part with nontrivial coefficients and the irrelevant part with small or zero coefficients. Meanwhile, we only assign independent hyperparameters to the coefficients in the significant part, while the remaining coefficients share a common hyperparameter. Then the generalized maximum likelihood (GML) criterion is adopted to classify the coefficients, ensuring both accuracy and efficiency. Hence, the parameter space in Bayesian inference will be significantly reduced, and the computational efficiency can be considerably promoted. Both theoretical analysis and numerical experiments validate that the proposed algorithm achieves superior performance with considerably improved computational efficiency in sample shortage scenarios.

Keywords: airborne radar; space-time adaptive processing; sparse Bayesian learning

1. Introduction

Moving target detection is a fundamental function of radar systems for military surveillance and reconnaissance. Space-time adaptive processing (STAP) has become an effective and mature clutter suppression technique for airborne early warning phased array radar systems [1–3]. The optimal STAP filter is constructed using the ideal clutter plus noise covariance matrix (CNCM). In practice, the ideal CNCM is unknown a priori and is typically estimated using independent and identically distributed (i.i.d.) samples around the cell under test (CUT). According to the well-known Reed–Mallet–Brennan criterion [4], the number of i.i.d. samples with at least twice the system degree of freedom (DOF) is required to ensure a performance loss under 3 dB. Unfortunately, sufficient i.i.d. samples are often challenging to obtain due to various terrains, artificial structures and array configurations.

Over the past few decades, many STAP methods have been developed to improve clutter suppression performance in heterogeneous environments, including classical data-dependent reduced-rank (RR) methods [5–8] and data-independent reduced-dimension (RD) methods [9–12]. Although the required number of training samples is reduced to twice the reduced dimension, the requirement is still difficult to meet under severe non-stationary clutter environments for modern large-scale radar systems.

During the recent fifteen years, many sparse recovery-based STAP (SR-STAP) algorithms have been proposed by exploiting the underlying connection between the compressed sensing technique and the intrinsic sparsity prior of the clutter spectrum [13–21].

With the help of sparse recovery theory, these SR-STAP algorithms have shown excellent clutter suppression performance with limited samples. Depending on the sparse recovery algorithms, SR-STAP can be generally classified into several categories, such as convex relaxation [13–15], greedy algorithms [16], Bayesian inference [17–19], and other methods [20,21].

Convex relaxation substitutes the ℓ_0 norm with the ℓ_1 norm as the sparse penalty. The ℓ_1 norm optimization problem has shown that sparse solutions can be stably obtained under certain conditions [22]. It has been extensively applied to the least absolute shrinkage and selection operator (LASSO) and basis pursuit (BP) based STAP methods [23]. Most convex relaxation methods require careful tuning of the regularization parameter, and inappropriate parameter selection will jeopardize the performance of clutter suppression as well as slow-moving target detection [14,15]. However, choosing an appropriate regularization parameter is quite challenging in practice.

In recent years, sparse Bayesian learning (SBL) has drawn much effort due to its preferable advantages, such as automatic self-regulation [18] and flexibility in exploiting the potential signal structure [24]. SBL was first proposed by Tipping in 2001 [25] and introduced to the field of STAP by Duan in 2017, termed M-SBL-STAP [17]. Numerous empirical results indicate that the SBL based SR-STAP can provide satisfactory performance and is quite robust to noise and high coherence dictionary [19]. However, M-SBL-STAP faces an overwhelming computational burden and large memory requirements, hindering its implementation in large-scale radar systems. Many efficient SBL based STAP algorithms have been developed to address this issue. In [18], Wang proposed a fast-converging SBL algorithm by combining an approximation term, but the global convergence property of the algorithm is not guaranteed. An iterative reweighted based M-SBL (M-SBL-IR$\ell_{2,1}$) STAP algorithm was proposed by Liu [19], which exhibits better reconstruction accuracy and has a favourable convergence speed.

Our experience with numerous simulation experiments demonstrates that the majority of the space-time coefficients to be recovered are zero or close to zero, and only a few have nontrivial values, owing to the sparse nature of the clutter. Therefore, in this paper, we are inspired to propose an improved Bayesian probabilistic model for SR-STAP, exploiting the aforementioned sparsity feature of the clutter. Instead of assigning a separate hyperparameter to each space-time coefficient in the conventional Bayesian model, only the significant coefficients are assigned independent hyperparameters, while the remaining irrelevant coefficients share a common hyperparameter. As a result, the parameter space to be updated will be dramatically reduced, and the computational efficiency will naturally be promoted. In [26], a heuristic expansion-compression variance-component based method (ExCoV) has been proposed to guide us to classify those hyperparameters.

Specifically, the main contributions of this paper are summarized as follows:

1. By exploiting the inherent sparsity nature of the clutter, the space-time coefficients are divided into two disjoint groups, i.e., the significant and irrelevant groups, thus reducing redundancy, scaling down the parameter space and yielding an improved Bayesian probabilistic model for SR-STAP with reduced computational complexity and reduced memory requirements.
2. The space-time coefficients are partitioned into the significant and irrelevant groups according to the generalized maximum likelihood (GML) criterion, which preserves both accuracy and efficiency, unlike the conventional SBL cost function.
3. We extend and modify the real-value ExCoV method to complex-value STAP applications to approximately maximize the GML objective function. Using the ExCoV scheme, it is unnecessary to specify convergence thresholds and maximum iteration times.
4. Extensive experiments as well as detailed comparative analyses are presented, such as clutter suppression performance and target detection performance, etc.

Notations used in this paper are as follows: vectors, matrices and scalars are denoted by bold lowercase, bold uppercase and italic letters, respectively. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$

stand for conjugate, transpose and conjugate transpose. $\otimes$ and $\odot$ are the Kronecker and Hadamard (elementwise) product. $trace(\cdot)$ is the trace operator. $\mathbb{C}$ represents the set of complex values. $\|\cdot\|_F$ and $\|\cdot\|_{2,0}$ are respectively defined as the Frobenius norm and $\ell_{2,0}$ mixed norm, which is the number of non-zero elements of the vector formed by the ℓ_2 norm of each row. $E[\cdot]$ denotes the expectation operator.

The remainder of the paper is organized as follows. In Section 2, the airborne radar signal model is established. The formulation of SR-STAP model is presented in Section 3. The proposed algorithm is introduced in Section 4. The computational complexity analysis is presented in Section 5. Experiments and analyses are carried out on Section 6. Finally, Section 7 discuss the conclusions.

2. Signal Model

In this research, assume that an airborne pulsed-Doppler early warning radar system with a side-looking uniform linear array (ULA), as depicted in Figure 1, has N omnidirectional elements with half wavelength interelement spacing $d = \lambda/2$ and transmits M coherent pulses over a coherent processing interval (CPI) at the fixed pulse repetition frequency (PRF) f_r. The aircraft platform is cruising with a constant velocity v_p.
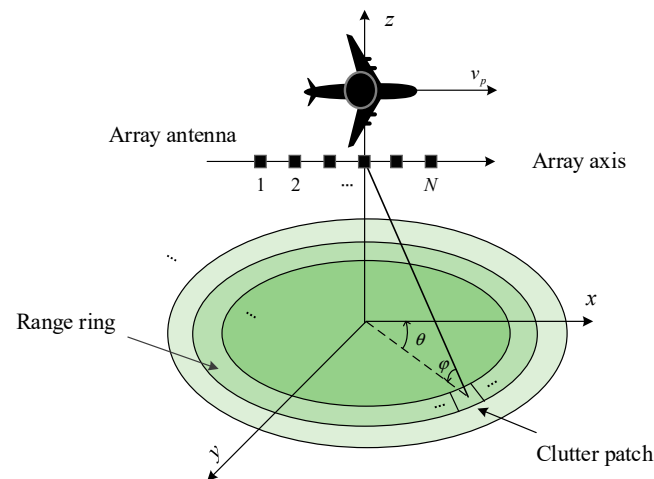

Figure 1. Geometric configuration of an airborne surveillance radar system.

Based on the well-known Ward clutter model [27], the ground can be divided into different range rings. The radar returns are composed of numerous evenly distributed and mutually uncorrelated clutter patches in azimuth angles θ. Considering the impact of range ambiguities, a general space-time snapshot $\mathbf{x} \in \mathbb{C}^{NM \times 1}$ from the CUT can be formulated as

$$
\begin{aligned}
\mathbf{x} &= \sum_{r=1}^{N_r} \sum_{i=1}^{N_c} \alpha_{r,i} \mathbf{s}(f_{d_{r,i}}, f_{s_{r,i}}) + \mathbf{n} \\
&= \sum_{r=1}^{N_r} \sum_{i=1}^{N_c} \alpha_{r,i} \left(\mathbf{s}_t(f_{d_{r,i}}) \otimes \mathbf{s}_s(f_{s_{r,i}}) \right) + \mathbf{n}
\end{aligned}
\tag{1}
$$

where $\alpha_{r,i}$ and $\mathbf{s}(f_{d_{r,i}}, f_{s_{r,i}})$ denote the complex amplitude and space-time steering vector from the corresponding clutter patch; N_c is the number of independent clutter patches in a iso-range ring; N_r represents the number of range ambiguities; the vector $\mathbf{n}$ is the thermal noise and modelled as a zero-mean complex Gaussian random process; $\mathbf{s}(f_{d_{r,i}}, f_{s_{r,i}})$, $\mathbf{s}_t(f_{d_{r,i}}) = \left[1, e^{j2\pi f_{d_{r,i}}}, \cdots, e^{j2\pi(M-1)f_{d_{r,i}}} \right]^T$ and $\mathbf{s}_s(f_{s_{r,i}}) = \left[1, e^{j2\pi f_{s_{r,i}}}, \cdots, e^{j2\pi(N-1)f_{s_{r,i}}} \right]^T$ stand for space-time, temporal and spatial steering vectors, respectively. According to the

geometric configuration of the radar platform shown in Figure 1, the normalized spatial frequency $f_{s_{r,i}}$ and normalized Doppler frequency $f_{d_{r,i}}$ are defined as follows:

$$f_{s_{r,i}} = \frac{d}{\lambda} \cos \varphi_r \cos \theta_i \tag{2}$$

$$f_{d_{r,i}} = \frac{2v_p}{\lambda f_r} \cos \varphi_r \cos \theta_i \tag{3}$$

where φ_r and θ_i are the elevation and azimuth angle of the (r, i)-th clutter patch, respectively.

Since the clutter patches are mutually independent, the ideal clutter plus noise covariance matrix (CNCM) of the CUT can be expressed as

$$\mathbf{R} = \sum_{r=1}^{N_r} \sum_{i=1}^{N_c} E\left\{ |\alpha_{r,i}|^2 \right\} \mathbf{s}(f_{d_{r,i}}, f_{s_{r,i}}) \mathbf{s}^H(f_{d_{r,i}}, f_{s_{r,i}}) + \sigma^2 \mathbf{I} \tag{4}$$

where σ^2 denotes the noise power.

Under the linearly constrained minimum variance (LCMV) criterion, the optimum STAP weight vector, which maximizes signal-to interference-plus-noise ratio (SINR), can be given by

$$\mathbf{w} = \frac{\mathbf{R}^{-1} \mathbf{s}_{target}}{\mathbf{s}_{target}^H \mathbf{R}^{-1} \mathbf{s}_{target}} \tag{5}$$

where $\mathbf{s}_{target}$ is the space-time steering vector of the desired target.

3. Sparse Recovery Based STAP Model Specification

For the grid-based SR-STAP algorithms [23], the whole continuous spatial-temporal plane is uniformly discretized into $K = N_s M_d$ grids, where $N_s = \rho_s N (\rho_s > 1)$ is the number of normalized spatial frequency bins in the spatial domain and $M_d = \rho_d M (\rho_d > 1)$ is the number of normalized Doppler frequency bins in the temporal domain. ρ_s and ρ_d are the resolution scales. Then $\boldsymbol{\Phi} = \left[\mathbf{s}(f_{d,1}, f_{s,1}), \cdots, \mathbf{s}(f_{d,M_d}, f_{s,N_s}) \right]$ is defined as the $NM \times K$ overcomplete dictionary matrix consisting of K grids. The SR-STAP signal model for the multiple measurement vectors (MMV) case $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{C}^{NM \times L}$ can next be reformulated as

$$\mathbf{X} = \boldsymbol{\Phi} \mathbf{A} + \mathbf{N} \tag{6}$$

where L is the number of IID space-time snapshots; $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_L] \in \mathbb{C}^{K \times L}$ denotes the space-time coefficients matrix where the non-zero rows indicate the potential presence of clutter components, and each column has the same sparsity support; $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_L] \in \mathbb{C}^{NM \times L}$ is the zero mean Gaussian white noise matrix.

Since $K \gg NM$, the problem of estimating $\mathbf{A}$ is fundamentally underdetermined. The canonical cost function of SR-STAP problem is formulated by

$$\min_{\mathbf{A}} \|\mathbf{X} - \boldsymbol{\Phi} \mathbf{A}\|_F^2 + \eta \|\mathbf{A}\|_{2,0} \tag{7}$$

where η is a nonnegative regularization parameter controlling the trade-off between the sparse penalty and data fidelity, and thus the choice of η is critical to the recovery performance. Unfortunately, finding the $\ell_{2,0}$ norm optimal representation requires a combinatorial search (also known as NP-hard [28]) and, therefore, is difficult to obtain.

Many recent alternative tractable approaches have been developed to find sparse solutions efficiently [13–21]. Motivated by the predominance of the SBL algorithms in SR-STAP, we resort to SBL for accurately recovering $\mathbf{A}$ from $\mathbf{X}$ in this research.

4. Proposed Algorithm

Following the conventional sparse Bayesian learning framework, all the unknowns are treated as stochastic variables with assigned independent probability distributions. First of

all, the noise matrix is modelled as white complex Gaussian noise with unknown power σ^2; thus the observed Gaussian likelihood function of the measurements $\mathbf{X}$ for the MMV case can be expressed as.

$$
\begin{aligned}
p(\mathbf{X}|\mathbf{A}, \sigma^2) &= \prod_{l=1}^{L} \mathcal{CN}\left(\mathbf{X}_{\cdot l} \big| \boldsymbol{\Phi}\mathbf{A}_{\cdot l}, \sigma^2 \mathbf{I}\right) \\
&= \frac{1}{(\pi\sigma^2)^{NML}} \exp\left[-\sigma^{-2}\|\mathbf{X} - \boldsymbol{\Phi}\mathbf{A}\|_F^2\right]
\end{aligned}
\tag{8}
$$

Since measurements are mutually independent and identically distributed, we suppose each column of the space-time coefficient matrix $\mathbf{A}_{\cdot l}$ is assigned with the same zero-mean complex Gaussian prior $\boldsymbol{\gamma}$, governing the prior variance of each unknown space-time coefficient.

$$
\begin{aligned}
p(\mathbf{A}|\boldsymbol{\gamma}) &= \prod_{l=1}^{L} \mathcal{CN}(\mathbf{A}_{\cdot l}|0, \boldsymbol{\Gamma}) \\
&= \pi^{-KL}|\boldsymbol{\Gamma}|^{-L} \exp\left(-\sum_{l=1}^{L} \mathbf{A}_{\cdot l}^H \boldsymbol{\Gamma}^{-1} \mathbf{A}_{\cdot l}\right)
\end{aligned}
\tag{9}
$$

where $0 \in \mathbb{C}^{K\times 1}$ is a zero vector, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_K]$ is the unknown variance hyperparameters corresponding to the rows in $\mathbf{A}$, and covariance matrix $\boldsymbol{\Gamma}$ is a diagonal matrix, where $\boldsymbol{\gamma}$ is its diagonal elements, i.e., $\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma})$.

Further, following the Bayesian theorem, the posterior probability density $p(\mathbf{A}|\mathbf{X}, \boldsymbol{\gamma}, \sigma^2)$ can be easily calculated by

$$
\begin{aligned}
p(\mathbf{A}|\mathbf{X}, \boldsymbol{\gamma}, \sigma^2) &= \frac{p(\mathbf{X}|\mathbf{A}, \sigma^2)p(\mathbf{A}|\boldsymbol{\gamma})}{p(\mathbf{X}|\boldsymbol{\gamma}, \sigma^2)} \\
&= \frac{p(\mathbf{X}|\mathbf{A}, \sigma^2)p(\mathbf{A}|\boldsymbol{\gamma})}{\int p(\mathbf{X}|\mathbf{A}, \sigma^2)p(\mathbf{A}|\boldsymbol{\gamma})d\mathbf{A}}
\end{aligned}
\tag{10}
$$

The posterior probability $p(\mathbf{A}|\mathbf{X}, \boldsymbol{\gamma}, \sigma^2)$ obeys a multivariate complex Gaussian distribution $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean and covariance respectively given by

$$
\boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\Phi}^H\left(\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H\right)^{-1}\mathbf{X}
\tag{11}
$$

$$
\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^H\left(\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H\right)^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}
\tag{12}
$$

Thus, with a fixed $\boldsymbol{\Gamma}$, the estimated sparse recovery solution $\widetilde{\mathbf{A}}$ of M-SBL is

$$
\widetilde{\mathbf{A}} = \boldsymbol{\Gamma}\boldsymbol{\Phi}^H\left(\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H\right)^{-1}\mathbf{X}
\tag{13}
$$

Accordingly, for $\gamma_i \to 0$, the corresponding row $\widetilde{\mathbf{A}}_{i\cdot}$ of the space-time coefficient matrix will be zeros as well. In other words, if $\boldsymbol{\gamma}$ is sparse, the corresponding space-time coefficient estimation $\widetilde{\mathbf{A}}$ will also be sparse.

The hyperparameters vector can be estimated by performing the type-II maximum likelihood procedure or evidence maximization in $\boldsymbol{\gamma}$ space [29]. Mathematically, the cost function can be expressed as minimizing the marginal likelihood function $\mathcal{L}(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$:

$$
\begin{aligned}
\widetilde{\boldsymbol{\gamma}} &= \arg\min_{\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\gamma}) \\
&= \arg\min_{\boldsymbol{\gamma}} -\ln p(\mathbf{X}|\boldsymbol{\gamma}, \sigma^2) \\
&= \arg\min_{\boldsymbol{\gamma}} -\ln \int p(\mathbf{X}|\mathbf{A}, \sigma^2)p(\mathbf{A}|\boldsymbol{\gamma})d\mathbf{A} \\
&= \arg\min_{\boldsymbol{\gamma}} L\ln|\mathbf{P}| + \sum_{l=1}^{L} \mathbf{x}_l^H \mathbf{P}^{-1} \mathbf{x}_l
\end{aligned}
\tag{14}
$$

After the unknown coefficients **A** have been integrated out, **P** stands for the covariance of the measurements **X** with the hyperparameters $\boldsymbol{\gamma}$ and σ^2.

$$\mathbf{P} \triangleq \sigma^2 \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Gamma} \boldsymbol{\Phi}^H \tag{15}$$

This minimization can be performed by an iterative reweighted SBL based algorithm [19,30], which will be modified and described in the following part of this paper.

As we can observe from (11) and (12), the computational bottleneck is mainly manifested in large-scale matrix multiplication and matrix inversion operations. The storage requirement is also heavy. Moreover, during each iteration, a hyperparameter space $(\boldsymbol{\gamma}, \sigma^2)$ of dimension $K + 1$ needs to be updated. These factors make SBL based STAP algorithms considerably slower than other types of sparse recovery algorithms, even though excellent clutter suppression performance can be achieved with limited training samples.

In fact, due to the intrinsic sparsity of the clutter spectrum on the spatial-temporal plane, only a few rows of **A** have nontrivial magnitudes, and the remaining elements are strictly zero (or close to zero). It implies that most of the hyperparameters in $\boldsymbol{\gamma}$ will converge to zeros, and they are redundant. Therefore, we can naturally consider partitioning the space-time coefficients into two parts: the significant elements part and the complementary irrelevant elements part. Instead of assigning independent hyperparameters to all coefficients in the conventional SBL framework, we can only allocate independent hyperparameters to significant coefficients part, while the remaining irrelevant coefficients share an identical hyperparameter. Consequently, the dimension of the hyperparameter space can be reduced to be proportional to the sparsity level of the clutter, which is much smaller than K [23]. As a result, the computation complexity will be greatly mitigated, and the memory consumption will be reduced, particularly in large-scale modern radar systems.

Next, define $\Theta = \{1, 2, \cdots, K\}$ as the complete index set. The index set of significant coefficients is denoted by Θ_α with size K_α, and the complementary index set of irrelevant coefficients is denoted by $\Theta_\beta = \Theta \backslash \Theta_\alpha$ with cardinality $K_\beta = K - K_\alpha$. Accordingly, we will also divide dictionary matrix $\boldsymbol{\Phi}$ and space-time coefficients matrix **A** into two submatrices, i.e., $\boldsymbol{\Phi}_\alpha \in \mathbb{C}^{NM \times K_\alpha}$, $\boldsymbol{\Phi}_\beta \in \mathbb{C}^{NM \times K_\beta}$, $\mathbf{A}_\alpha \in \mathbb{C}^{K_\alpha \times L}$ and $\mathbf{A}_\beta \in \mathbb{C}^{K_\beta \times L}$, respectively. For more details:

● $\boldsymbol{\Phi}_\alpha$ is the submatrix of $\boldsymbol{\Phi}$ corresponding to the significant coefficients index set Θ_α, e.g., if $\Theta_\alpha = \{2, 8, 9\}$, then $\boldsymbol{\Phi}_\alpha = [\mathbf{s}_2, \mathbf{s}_8, \mathbf{s}_9]$, where $\mathbf{s}_i$ is the ith column of $\boldsymbol{\Phi}$.

● $\mathbf{A}_\alpha$ is the submatrix of **A** corresponding to the index set Θ_α, e.g., if $\Theta_\alpha = \{2, 8, 9\}$, then $\mathbf{A}_\alpha = [\mathbf{A}_{2.} ; \mathbf{A}_{8.} ; \mathbf{A}_{9.}]$, where $\mathbf{A}_{i.}$ is the ith row of **A**.

Then (9) can be reformulated as

$$\begin{aligned} p(\mathbf{A} \big| \boldsymbol{\gamma}) &= p(\mathbf{A}_\alpha \big| \boldsymbol{\gamma}_{\Theta_\alpha}; \Theta_\alpha) p(\mathbf{A}_\beta \big| \boldsymbol{\gamma}_{\Theta_\beta}; \Theta_\beta) \\ &= \prod_{l=1}^{L} \Big(\mathcal{CN}(\mathbf{A}_{\alpha.l} \big| 0, \boldsymbol{\Gamma}_\alpha) \cdot \mathcal{CN}(\mathbf{A}_{\beta.l} \big| 0, \gamma_{\Theta_\beta} \mathbf{I}) \Big) \\ &= \pi^{-KL} |\boldsymbol{\Gamma}_\alpha|^{-L} (\gamma_{\Theta_\beta})^{-LK_\beta} \\ &\quad \cdot \exp\Big(-\sum_{l=1}^{L} \Big(\mathbf{A}_{\alpha l}^H \boldsymbol{\Gamma}_\alpha^{-1} \mathbf{A}_{\alpha l} + \gamma_{\Theta_\beta}^{-1} \mathbf{A}_{\beta l}^H \mathbf{A}_{\beta l} \Big) \Big) \end{aligned} \tag{16}$$

where $\boldsymbol{\Gamma}_\alpha$ and $\gamma_{\Theta_\beta} \mathbf{I}$ denote the covariance matrix of the significant part $\mathbf{A}_\alpha$ and irrelevant part $\mathbf{A}_\beta$, respectively.

If all the coefficients belong to the Θ_α index set and $K_\alpha = K$, then the above probabilistic model will degenerate to the original SBL full model in (9). Similarly, according to the index sets Θ_α and Θ_β, (11) and (12) can be respectively partitioned into submatrices

$$\boldsymbol{\mu}_\alpha = \boldsymbol{\Gamma}_\alpha \boldsymbol{\Phi}_\alpha^H \mathbf{P}^{-1} \mathbf{X} \tag{17}$$

$$\boldsymbol{\mu}_\beta = \gamma_{\Theta_\beta} \boldsymbol{\Phi}_\beta^H \mathbf{P}^{-1} \mathbf{X} \tag{18}$$

$$\boldsymbol{\Sigma}_\alpha = \boldsymbol{\Gamma}_\alpha - \boldsymbol{\Gamma}_\alpha \boldsymbol{\Phi}_\alpha^H \mathbf{P}^{-1} \boldsymbol{\Phi}_\alpha^H \boldsymbol{\Gamma}_\alpha \tag{19}$$

$$\mathbf{\Sigma}_\beta = \gamma_{\Theta_\beta} \mathbf{I}_{K_\beta} - \gamma_{\Theta_\beta}^2 \mathbf{\Phi}_\beta^H \mathbf{P}^{-1} \mathbf{\Phi}_\beta^H \tag{20}$$

where the covariance matrix $\mathbf{P}$ in (15) now can be rewritten as

$$\mathbf{P} = \sigma^2 \mathbf{I} + \mathbf{\Phi}_\alpha \mathbf{\Gamma}_\alpha \mathbf{\Phi}_\alpha^H + \gamma_{\Theta_\beta} \mathbf{\Phi}_\beta \mathbf{\Phi}_\beta^H \tag{21}$$

For the sake of expression simplicity, we define the set of all unknowns:

$$\mathbf{\psi} = (\mathbf{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}, \sigma^2, \Theta_\alpha) \tag{22}$$

Now, let us consider estimating the most superior and efficient coefficient index sets Θ_α and Θ_β, i.e., the optimal representation. According to [26], the generalized maximum-likelihood (GML) criterion can be used to assign the hyperparameters, and its equivalence with the canonical optimization problem in (7) is demonstrated. The GML criterion maximizes:

$$\text{GML}(\mathbf{\psi}) = \ln p(\mathbf{X}|\mathbf{\psi}) - \ln|\mathcal{I}(\mathbf{\psi})| \tag{23}$$

where the first term is the same marginal likelihood function in (14) with Θ_α fixed, enforcing the estimate fit the measurements; the second term $\mathcal{I}(\mathbf{\psi})$ is the Fisher information matrix (FIM) for the hyperparameters. Based on the well-known FIM for the Gaussian measurement model [31], we can deduce the block-partitioned FIM result for the hyperparameters vector (the detailed derivation is shown in Appendix A).

$$\mathcal{I}(\mathbf{\psi}) = \begin{bmatrix} \mathcal{I}_{\mathbf{\gamma}_{\Theta_\alpha}, \mathbf{\gamma}_{\Theta_\alpha}}(\mathbf{\psi}) & \mathcal{I}_{\mathbf{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}}(\mathbf{\psi}) \\ \mathcal{I}_{\mathbf{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}}^H(\mathbf{\psi}) & \mathcal{I}_{\gamma_{\Theta_\beta}, \gamma_{\Theta_\beta}}(\mathbf{\psi}) \end{bmatrix} \tag{24}$$

with each block computed as

$$\mathcal{I}_{\mathbf{\gamma}_{\Theta_\alpha}, \mathbf{\gamma}_{\Theta_\alpha}}(\mathbf{\psi}) = 2L \left(\mathbf{\Phi}_\alpha^H \mathbf{P}^{-1} \mathbf{\Phi}_\alpha \right) \odot \left(\mathbf{\Phi}_\alpha^H \mathbf{P}^{-1} \mathbf{\Phi}_\alpha \right) \tag{25}$$

$$\left[\mathcal{I}_{\mathbf{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}}(\mathbf{\psi}) \right]_i = 2L \mathbf{\Phi}_\alpha^H(:, i) \mathbf{P}^{-1} \mathbf{\Phi}_\beta \mathbf{\Phi}_\beta^H \mathbf{P}^{-1} \mathbf{\Phi}_\alpha(:, i) \tag{26}$$

$$\mathcal{I}_{\gamma_{\Theta_\beta}, \gamma_{\Theta_\beta}}(\mathbf{\psi}) = 2L \cdot trace(\mathbf{P}^{-1} \mathbf{\Phi}_\beta \mathbf{\Phi}_\beta^H \mathbf{P}^{-1} \mathbf{\Phi}_\beta \mathbf{\Phi}_\beta^H) \tag{27}$$

Since the hyperparameter space is nested, namely any Θ_α is a subset of Θ, the more parameters in the set Θ_α, the larger the value $\ln p(\mathbf{X}|\mathbf{\psi})$ will be, which reduces the model mismatch. Therefore, there will always be a tendency to choose the complete set $\Theta_\alpha = \Theta$ [31]. However, the second term increases as the set Θ_α grows, penalising the growth of the set. The GML criterion is, therefore, able to maintain a balance between underfitting and overfitting of the model, ensuring both accuracy and efficiency.

Nevertheless, directly applying the GML criterion (23) to obtain the optimal parameter index sets still requires an exhaustive search and is hard to implement in practical applications. Subsequently, we will employ an ExCoV based method to maximize the objective function approximately [32]. The fundamental idea is interleaving the expansion/compression step and the parameter update step. In each expansion or compression procedure, we modify the current estimate of Θ_α by one element per step to obtain a larger GML($\mathbf{\psi}$). Then followed by the parameter update step, the hyperparameters $(\mathbf{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}, \sigma^2)$ are updated for a fixed index set Θ_α. Subsequently, we introduce the three main iteration steps involved in ExCoV: expansion step, compression step and parameter update step.

A. Expansion step. In this step, we determine the index $q \in \Theta_\beta^{(p)}$ corresponding to the row of $\mathbf{\mu}_\beta^{(p)}$ with the largest ℓ_2 norm

$$q = \arg \max_{q \in \mathbf{\Theta}_\beta^{(p)}} \| \left(\mathbf{\mu}_\beta^{(p)} \right)_{q \cdot} \|_2 \tag{28}$$

where the superscript $(\cdot)^{(p)}$ represents the iteration number. Then the index q is moved from $\Theta_\beta{}^{(p)}$ to $\Theta_\alpha{}^{(p)}$, i.e.,$\Theta_\alpha{}^{(p+1)} = \Theta_\alpha{}^{(p)} \cup \{q\}$ and $\Theta_\beta{}^{(p+1)} = \Theta_\beta{}^{(p)} \backslash \{q\}$, yielding $K_\alpha{}^{(p+1)} = K_\alpha{}^{(p)} + 1$, $K_\beta{}^{(p+1)} = K_\beta{}^{(p)} - 1$. The new expanded vector $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)}$ can be expressed as $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)} = \left[\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}}{}^{(p)}; \gamma_{\Theta_\beta}{}^{(p)} \right]$.

B. Compression step. Here, we determine the index $t \in \Theta_\alpha{}^{(p)}$ corresponding to the smallest element of $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}}{}^{(p)}$

$$t = \arg \min_{t \in \Theta_\alpha{}^{(p)}} \left(\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}, t} \right)^{(p)} \tag{29}$$

Then the index t is moved from $\Theta_\alpha{}^{(p)}$ to $\Theta_\beta{}^{(p)}$, i.e., $\Theta_\beta{}^{(p+1)} = \Theta_\beta{}^{(p)} \cup \{t\}$ and $\Theta_\alpha{}^{(p+1)} = \Theta_\alpha{}^{(p)} \backslash \{t\}$, yielding $K_\alpha{}^{(p+1)} = K_\alpha{}^{(p)} - 1$, $K_\beta{}^{(p+1)} = K_\beta{}^{(p)} + 1$. The new compressed vector $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)}$ can be constructed as $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)} = \left[\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}, 1}{}^{(p)}, \cdots, \boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}, t-1}{}^{(p)}, \boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}, t+1}{}^{(p)}, \cdots, \boldsymbol{\gamma}_{\Theta_\alpha{}^{(p)}, K_\alpha{}^{(p)}}{}^{(p)} \right]$.

C. Parameters update step. In [26], the original ExCoV algorithm employs the expectation-maximization (EM) method for Bayesian inference to estimate the hyperparameters, which usually requires extensive iterations to converge [18]. To overcome this drawback and thus further improve code running speed, an iterative reweighted M-SBL strategy is modified to update hyperparameters.

In this step, we firstly assume that the index sets Θ_α and Θ_β are known and fixed. We next present the derivation of the parameters update step based on the iterative reweighed $\ell_{2,1}$ M-SBL algorithm from [19] modified by the previous proposed Bayesian probabilistic model (details can be seen in Appendix B), which can remarkably accelerate the convergence speed. Then, the hyperparameters $\left(\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p+1)}, \gamma_{\Theta_\beta}{}^{(p+1)}, (\sigma^2)^{(p+1)} \right)$ can be updated by

$$\boldsymbol{\gamma}_{\Theta_{\alpha,i}{}^{(p+1)}}{}^{(p+1)} = \left(\frac{\frac{1}{L} \sum_{l=1}^{L} \left| \boldsymbol{\mu}_{\alpha il}{}^{(p+1)} \right|^2}{\mathbf{s}_i^H \left(\mathbf{P}^{(p+1)} \right)^{-1} \mathbf{s}_i} \right)^{\frac{1}{2}} \tag{30}$$

$$\gamma_{\Theta_\beta}{}^{(p+1)} = \left(\frac{\frac{1}{L} \sum_{j=1}^{K_\beta} \sum_{l=1}^{L} \left| \boldsymbol{\mu}_{\beta jl}{}^{(p+1)} \right|^2}{K_\beta \cdot trace \left(\left(\mathbf{P}^{(p+1)} \right)^{-1} \boldsymbol{\Phi}_\beta \boldsymbol{\Phi}_\beta^H \right)} \right)^{\frac{1}{2}} \tag{31}$$

$$(\sigma^2)^{(p+1)} = \frac{\|\mathbf{X} - \boldsymbol{\Phi}\boldsymbol{\mu}\|_F^2 + L(\sigma^2)^{(p)} \left(N - (\sigma^2)^{(p)} \cdot trace \left((\mathbf{P}^{(p+1)})^{-1} \right) \right)}{NML} \tag{32}$$

where $\mathbf{P}^{(p+1)} = \left((\sigma^2)^{(p)} \mathbf{I} + \boldsymbol{\Phi}_\alpha \boldsymbol{\Gamma}_\alpha^{(p)} \boldsymbol{\Phi}_\alpha^H + \gamma_{\Theta_\beta}{}^{(p)} \boldsymbol{\Phi}_\beta \boldsymbol{\Phi}_\beta^H \right)$.

We term the proposed algorithm as an improved iterative reweighted sparse Bayesian learning algorithm based on expansion-compression variance components (ExCoV-IIR-MSBL). Figure 2 illustrates the processing flowchart of the proposed algorithm and its procedures are summarized as follows.
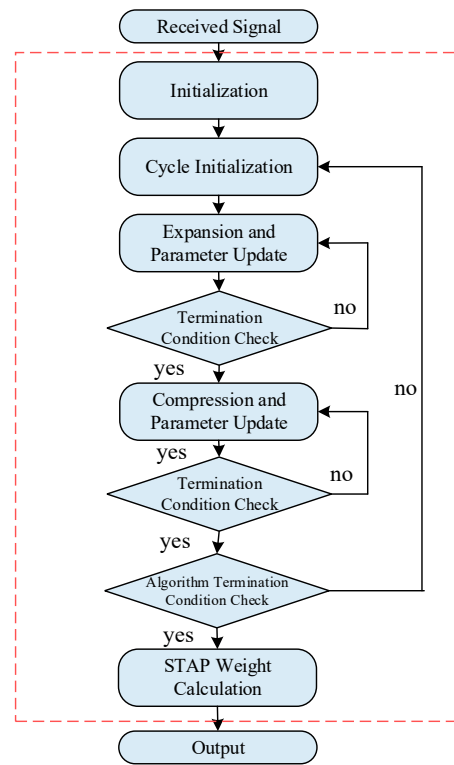
Figure 2. Processing flowchart of the proposed algorithm.

Step1 Initialization: Calculate the minimum norm solution of the space-time coefficients

$$\mathbf{\mu}^{(0)} = \mathbf{\Phi}^H \left(\mathbf{\Phi}^H \mathbf{\Phi} \right)^{-1} \mathbf{X} \tag{33}$$

Then we can utilize the prior knowledge provided by the inertial navigation system and radar system to initialize $K_\alpha^{(0)}$

$$K_\alpha^{(0)} = \lfloor N + \beta(M-1) \rfloor \tag{34}$$

where $\beta = 2v_p / (f_r d)$ represents the slope of the clutter ridge [33] and $\lfloor \cdot \rfloor$ denotes rounding-up operation; or simply set $K_\alpha^{(0)} = 1$, since the ExCoV method is insensitive to the initial value of K_a. Hence, the index set $\Theta_\alpha^{(0)}$ is constructed using the rows of $\mathbf{\mu}^{(0)}$ corresponding to the first $K_\alpha^{(0)}$ largest ℓ_2 norm. Subsequently, $K_\beta^{(0)} = K - K_\alpha^{(0)}$ and $\Theta_\beta^{(0)} = \Theta \backslash \Theta_\alpha^{(0)}$.

Step2 Cycle Initialization: Set the initial values for the hyperparameters $(\mathbf{\gamma}_{\Theta_\alpha}, \mathbf{\gamma}_{\Theta_\beta}, \sigma^2)$.

$$\left(\sigma^2 \right)^{(0)} = \frac{1}{NML} \| \mathbf{X} - \mathbf{\Phi} \mathbf{A} \|_F^2 \tag{35}$$

$$\mathbf{\gamma}_{\Theta_\alpha, i}^{(0)} = \frac{1}{L} \sum_{l=1}^{L} \frac{\left| \mathbf{s}_i^H \mathbf{x}_l \right|^2}{\left| \mathbf{s}_i^H \mathbf{s}_i \right|^2}, \quad \left(\mathbf{s}_i \in \mathbf{\Phi}_\alpha \ and \ i = 1, \cdots, K_\alpha^{(0)} \right) \tag{36}$$

$$\gamma_{\Theta_\beta}{}^{(0)} = \min_i \left(\mathbf{\gamma}_{\Theta_\alpha, i} \right)^{(0)} \tag{37}$$

Generally, these parameters can be initialized to an arbitrary positive vector, but it would be more beneficial if a rough estimate could be given. Then the initial values of the optimal record are set to $\hat{\mathbf{\psi}} = \left(\mathbf{\gamma}_{\Theta_\alpha}^{(0)}, \gamma_{\Theta_\beta}{}^{(0)}, \left(\sigma^2 \right)^{(0)}, \Theta_\alpha^{(0)} \right)$ and $\hat{\mathbf{\mu}} = \mathbf{\mu}^{(0)}$.

Step3 Expansion: Apply the aforementioned expansion step, yielding the updated $\Theta_\alpha{}^{(p+1)}, \Theta_\beta{}^{(p+1)}$ and $\mathbf{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)}$.

Step4 Parameter Update: With the aforementioned modified iterative reweighted M-SBL update step, the updated parameters set $\boldsymbol{\psi}^{(p+1)} = \left(\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p+1)}, \boldsymbol{\gamma}_{\Theta_\beta{}^{(p+1)}}{}^{(p+1)}, (\sigma^2)^{(p+1)}, \Theta_\alpha{}^{(p+1)}\right)$ and space-time coefficients estimate $\boldsymbol{\mu}^{(p+1)}$ are obtained.

Step5 Optimal Estimates Update: During the iterations, we shall record the optimal unknown parameters set $\hat{\boldsymbol{\psi}}$ that maximizes the GML, and the corresponding space-time coefficients estimates $\hat{\boldsymbol{\mu}}_\alpha$ and $\hat{\boldsymbol{\mu}}_\beta$ in the latest iteration cycle. After the unknown parameters $\boldsymbol{\psi}^{(p+1)} = \left(\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p+1)}, \boldsymbol{\gamma}_{\Theta_\beta{}^{(p+1)}}{}^{(p+1)}, (\sigma^2)^{(p+1)}, \Theta_\alpha{}^{(p+1)}\right)$ are updated, the new GML$(\boldsymbol{\psi}^{(p+1)})$ can be calculated using (23). Next, we will verify whether the following condition holds

$$\text{GML}\left(\boldsymbol{\psi}^{(p+1)}\right) > \text{GML}\left(\hat{\boldsymbol{\psi}}\right) \tag{38}$$

If the inequality holds, then we set $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(p+1)}$, $\hat{\boldsymbol{\mu}}_\alpha = \boldsymbol{\mu}_\alpha{}^{(p+1)}$ and $\hat{\boldsymbol{\mu}}_\beta = \boldsymbol{\mu}_\beta{}^{(p+1)}$; instead, keep $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\mu}}_\alpha$ and $\hat{\boldsymbol{\mu}}_\beta$ unchanged.

Step6 Termination Condition Check: In this step, we would verify whether the expansion is still required to continue [26].

$$\text{GML}\left(\boldsymbol{\psi}^{(p+1)}\right) < \min\left\{\text{GML}\left(\boldsymbol{\psi}^{(p+1-T)}\right), \frac{1}{T}\sum_{t=0}^{T-1}\text{GML}\left(\boldsymbol{\psi}^{(p-t)}\right)\right\} \tag{39}$$

where T is the length of a sliding average window. This condition will help us to determine if there is still a need for the expansion operation and prevents premature termination. If the inequality is satisfied, then the expansion operation is terminated, and Step7 begins; otherwise, the expansion continues and returns to Step3.

Step7 Compression: Apply the aforementioned compression step, yielding the updated $\Theta_\alpha{}^{(p+1)}$, $\Theta_\beta{}^{(p+1)}$ and $\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p)}$.

Step8 Parameter Update: With the aforementioned iterative reweighted M-SBL update step, the updated parameters set $\boldsymbol{\psi}^{(p+1)} = \left(\boldsymbol{\gamma}_{\Theta_\alpha{}^{(p+1)}}{}^{(p+1)}, \boldsymbol{\gamma}_{\Theta_\beta{}^{(p+1)}}{}^{(p+1)}, (\sigma^2)^{(p+1)}, \Theta_\alpha{}^{(p+1)}\right)$ and space-time coefficients estimate $\boldsymbol{\mu}^{(p+1)}$ are obtained.

Step9 Optimal Estimates Update: Verify whether the condition (38) holds. If the inequality holds, then we set $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(p+1)}$, $\hat{\boldsymbol{\mu}}_\alpha = \boldsymbol{\mu}_\alpha{}^{(p+1)}$ and $\hat{\boldsymbol{\mu}}_\beta = \boldsymbol{\mu}_\beta{}^{(p+1)}$; instead, keep $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\mu}}_\alpha$ and $\hat{\boldsymbol{\mu}}_\beta$ unchanged.

Step10 Termination Condition Check: In this step, we would verify whether the compression is still required to continue and check the same condition (39) in Step6. If the inequality is satisfied, then the compression operation is terminated, and Step11 begins; otherwise, the compression continues and returns to Step7.

Step11 Globally Optimal Estimates Update: Moreover, we keep a record of the globally optimal parameter set $\hat{\boldsymbol{\psi}}^\star$ and the corresponding coefficients $\hat{\boldsymbol{\mu}}^\star$ in the entire iteration cycles, verifying the condition

$$\text{GML}(\hat{\boldsymbol{\psi}}) > \text{GML}\left(\hat{\boldsymbol{\psi}}^\star\right) \tag{40}$$

If the inequality holds, then the globally optimal estimates $\hat{\boldsymbol{\psi}}^\star$ and $\hat{\boldsymbol{\mu}}^\star$ can be updated with $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\mu}}$; instead, keep them unchanged.

Step12 Algorithm Iteration Termination Check: If the globally optimal index set $\hat{\Theta}_\alpha{}^\star$ is consistent between two consecutive cycles. The proposed algorithm is next terminated and the final globally optimal space-time coefficients $\hat{\boldsymbol{\mu}}^\star$ and noise power $(\sigma^2)^\star$ are output.

Otherwise, the globally optimal index set $\hat{\Theta}_\alpha{}^\star$ is updated, then the algorithm continues. We reset the iteration index $p = 0$, $K_\alpha^{(0)} = \hat{K}_\alpha{}^\star$, and calculate the initial values for the space-time coefficients

$$\boldsymbol{\mu}^{(0)} = \hat{\boldsymbol{\mu}}^\star + \boldsymbol{\Phi}^H \left(\boldsymbol{\Phi}^H \boldsymbol{\Phi}\right)^{-1}\left(\mathbf{X} - \boldsymbol{\Phi}\hat{\boldsymbol{\mu}}^\star\right) \tag{41}$$

Step13 STAP Weight Calculation: Once the estimated sparse recovery solution $\widetilde{\mathbf{A}} = \hat{\boldsymbol{\mu}}^\star$ and $\widetilde{\sigma}^2 = \left(\sigma^2\right)^\star$ is obtained, the CNCM can be reconstructed from

$$\widetilde{\mathbf{R}} = \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{K}\left|\widetilde{\mathbf{A}}_{il}\right|^2 \mathbf{s}_i \mathbf{s}_i^H + \widetilde{\sigma}^2 \mathbf{I}_{NM} \tag{42}$$

Then the STAP weight $\mathbf{w}$ can be obtained by using (5).

Step14 Output: Give the filtered output of the cell under test $y = \mathbf{w}^H \mathbf{x}_{\text{CUT}}$.

From the above steps, it is worth noting that the proposed algorithm is fully automatic and does not require setting any convergence threshold as well as the maximum iteration number.

5. Computational Complexity Analysis

In this Section, we will analyse the computational complexity of the proposed ExCoV-IIR-MSBL algorithm and compare it with some state-of-the-art SR-STAP algorithms, including M-CVX [34], M-OMP [35], M-FOCUSS [36], M-IAA [21], M-SBL [17], M-FCSBL [18] and M-SBL-IR$\ell_{2,1}$ [19]. Their computational complexity would be measured by the number of complex multiplications in a single iteration. The results are listed in Table 1, where r_s stands for the clutter sparsity level.

Table 1. Computational complexity comparison.

Algorithm	Computational Complexity for a Single Iteration
M-CVX	$O\left((KL)^3\right)$
M-OMP	$O\left(r_s^3 + r_s^2 NM + 2r_s NM + NMKL\right)$
M-FOCUSS	$O\left(2(NM)^2 K + (NM)^3 + NMKL\right)$
M-IAA	$O\left(2(NM)^2 K + (NM)^3 + NMKL\right)$
M-SBL	$O\left(3K^2 NM + 2(NM)^2 K + (NM)^3 + NMKL\right)$
M-FCSBL	$O\left(\begin{array}{c} 5(NM)^2 K + (NM)^3 + 3K + L \\ (2KL + 4K + L)NM \end{array}\right)$
M-SBL-IR$\ell_{2,1}$	$O\left(3K^2 NM + 3(NM)^2 K + (NM)^3 + NMKL\right)$
ExCoV-IIR-MSBL	$O\left(\begin{array}{c} K_\alpha^2 NM + 2(NM)^2 K + 2(NM)^3 + \\ (NM)^2 L + NMKL \end{array}\right)$

In the conventional sparse Bayesian learning probabilistic model, the computational load is greatly centred on the parameter update step, see (11) and (12). The computational complexity is in the order of $O\left(3K^2 NM + 2(NM)^2 K + (NM)^3 + NMKL\right)$, with large-scale matrix multiplication and matrix inversion dominating a major part. Moreover, the memory requirement is $O\left(K^2\right)$. Therefore, these drawbacks make the MSBL-STAP challenging to implement in modern large-scale airborne radar systems, despite achieving an excellent clutter suppression performance with only a few snapshots.

The primary computational burden of the proposed algorithm is concentrated on the parameter update and GML criterion calculation steps. Since the parameter space dimension is significantly reduced, the scale of matrix multiplication is also significantly decreased, given that $K_\alpha \ll K$. In addition, unlike other SBL algorithms, the proposed algorithm no longer needs to update the variance matrix (12); hence, the computational complexity can be further reduced. As a result, the computational complexity is in the

order of $O\left(K_\alpha^2 NM + 2(NM)^2 K + 2(NM)^3 + (NM)^2 L + NMKL\right)$, and the required memory consumption is $O(NMK)$. It can be seen that the computational load and memory consumption of the proposed algorithm have been remarkably reduced. These advantages make the proposed algorithm more suitable for modern large-scale airborne radar systems.

For a more intuitive comparison of the above-mentioned algorithms, we plot Figure 3, illustrating the relationship between the number of complex multiplications and the number of pulses. We assume that $\rho_s = \rho_d = 4$, $N = 8$, $L = 8$, and K_α is set to be the rank of clutter. It can be seen from Figure 3 that the M-SBL-IR$\ell_{2,1}$ has nearly the same computational load as M-SBL, and the proposed algorithm has the lowest computational complexity among these compared SBL based algorithms and is close to the M-IAA and M-FOCUSS algorithms.
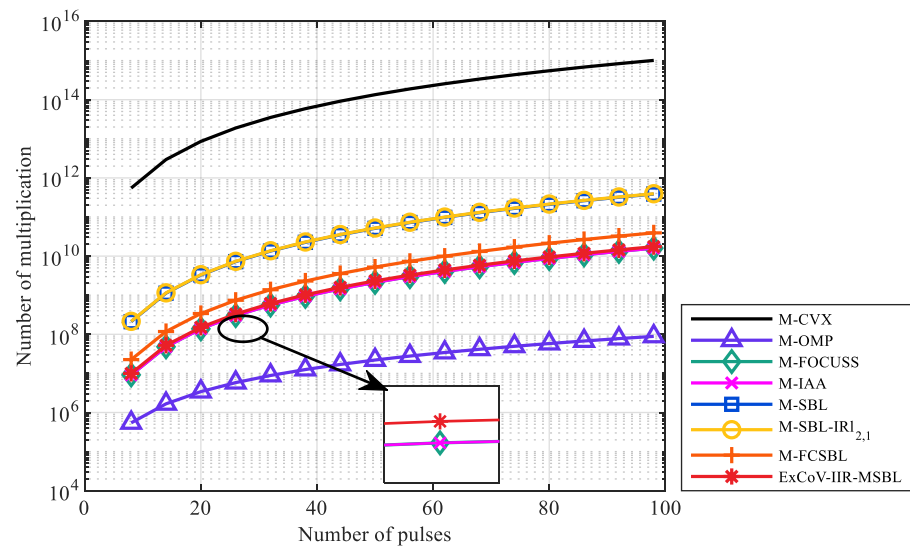
Figure 3. Computational complexity of different compared SR-STAP algorithms.

6. Numerical Experiment

In this section, numerous experiments are performed with simulated data to evaluate the performance of the proposed algorithm. The simulated data are generated by the well-known Ward clutter model introduced in the previous section. The main simulation parameters of a side-looking ULA are listed in Table 2. The dictionary resolution scales are set to be $\rho_s = \rho_d = 4$. Furthermore, all the simulation results are averaged over 100 independent Monte-Carlo trails.

Table 2. Simulation Parameters.

Parameter	Value
Array element number N	8
Pulse number M	8
PRF f_r	2000 Hz
Wavelength λ	0.3 m
Bandwidth B	2.5 MHz
Platform velocity v_p	150 m/s
CNR	50 dB
Clutter patch number N_c	1801

First of all, we examine the average code running time of the proposed algorithm and compare it with other state-of-the-art SR-STAP algorithms, as listed in Table 3. M-CVX, M-IAA, M-OMP, M-FOCUSS, M-SBL, M-FCSBL and M-SBL-IR$\ell_{2,1}$ are employed as benchmarks. Note that all the simulations are operated on the same workstation with Intel Xeon 4114 CPU @2.2GHz and 128 GB RAM.

Table 3. Running Time Comparison.

| Algorithm	Average Running Time
M-CVX	94.06 s
M-SBL	11.75 s
M-FOCUSS	10.48 s
M-SBL-IR$\ell_{2,1}$	6.16 s
M-FCSBL	2.20 s
M-IAA	2.15 s
ExCoV-IIR-MSBL	0.57 s
M-OMP	0.01 s

As seen from Table 3, since the parameter space is dramatically shrunk, the average running time of the proposed ExCoV-IIR-SBL is remarkably faster than the other compared SBL based STAP algorithms. It is even better than M-FOCUSS and M-IAA since the proposed algorithm can reach the steady state with fewer iterations with the help of a modified iterative reweighed parameter update step. Thus, the efficiency of the proposed algorithm is demonstrated.

In the following experiments, we assess the clutter suppression performance via the metric of signal to interference plus noise ratio (SINR) loss [27], defined as the ratio of output SINR to the optimum output SNR in a noise-only environment, i.e.,

$$SINR_{loss} = \frac{\sigma^2}{NM} \frac{\left| \mathbf{w}^H \mathbf{s}(f_d, f_s) \right|^2}{\mathbf{w}^H \mathbf{R} \mathbf{w}} \tag{43}$$

where $\mathbf{R}$ is the clairvoyant CNCM of the CUT, and $\mathbf{w}$ is the STAP weight.

Next, the clutter suppression performance in the ideal case is considered. The SINR loss performance of the proposed algorithm is evaluated and compared with other state-of-the-art SR-STAP algorithms, including M-CVX, M-IAA, M-OMP, M-FOCUSS, M-SBL, M-FCSBL and M-SBL-IR$\ell_{2,1}$. The number of i.i.d. training samples are set to be $L = 9$. Figure 4 depicts the curves of the SINR loss against the normalized Doppler frequency. As clearly illustrated in Figure 4, the proposed algorithm achieves the same near-optimal performance as M-IAA and the other three SBL based algorithms, revealing that the novel Bayesian probabilistic model proposed in this paper for STAP can accurately reconstruct the CNCM with a smaller parameter space. The performance of M-FOCUSS is found to be slightly inferior to theirs. Meanwhile, the proposed ExCoV-IIR-MSBL consumes the least running time among these three SBL based algorithms on this basis.
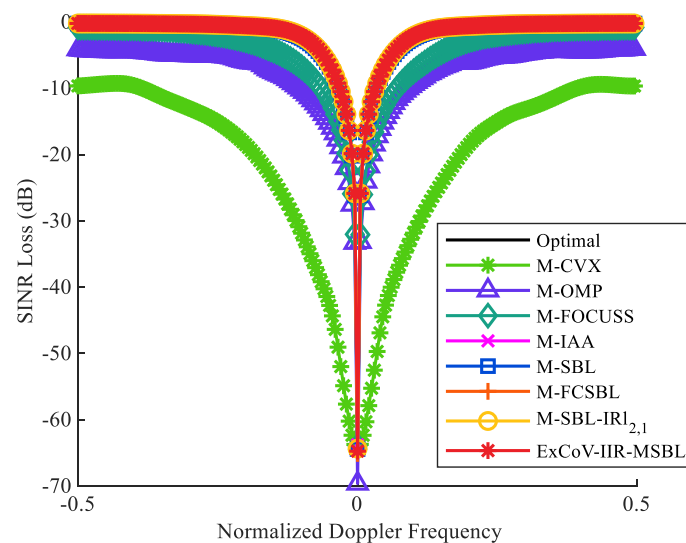
Figure 4. SINR loss comparison of different algorithms in the ideal case.

To more explicitly demonstrate the clutter suppression performance, Figure 5 plots the spatial-temporal adapted responses, given the STAP weight formed by the abovementioned algorithms. The presumed main lobe is located at a normalized spatial frequency of 0 with a normalized Doppler frequency of 0.2. From Figure 5, it is evident that the proposed method is able to maintain high gain at the presumed target location while precisely forming notches at the clutter ridge, suppressing both main lobe and sidelobe of the clutter component. Poorly performing algorithms have distorted two-dimensional responses and cannot accurately form notches at the clutter ridge.
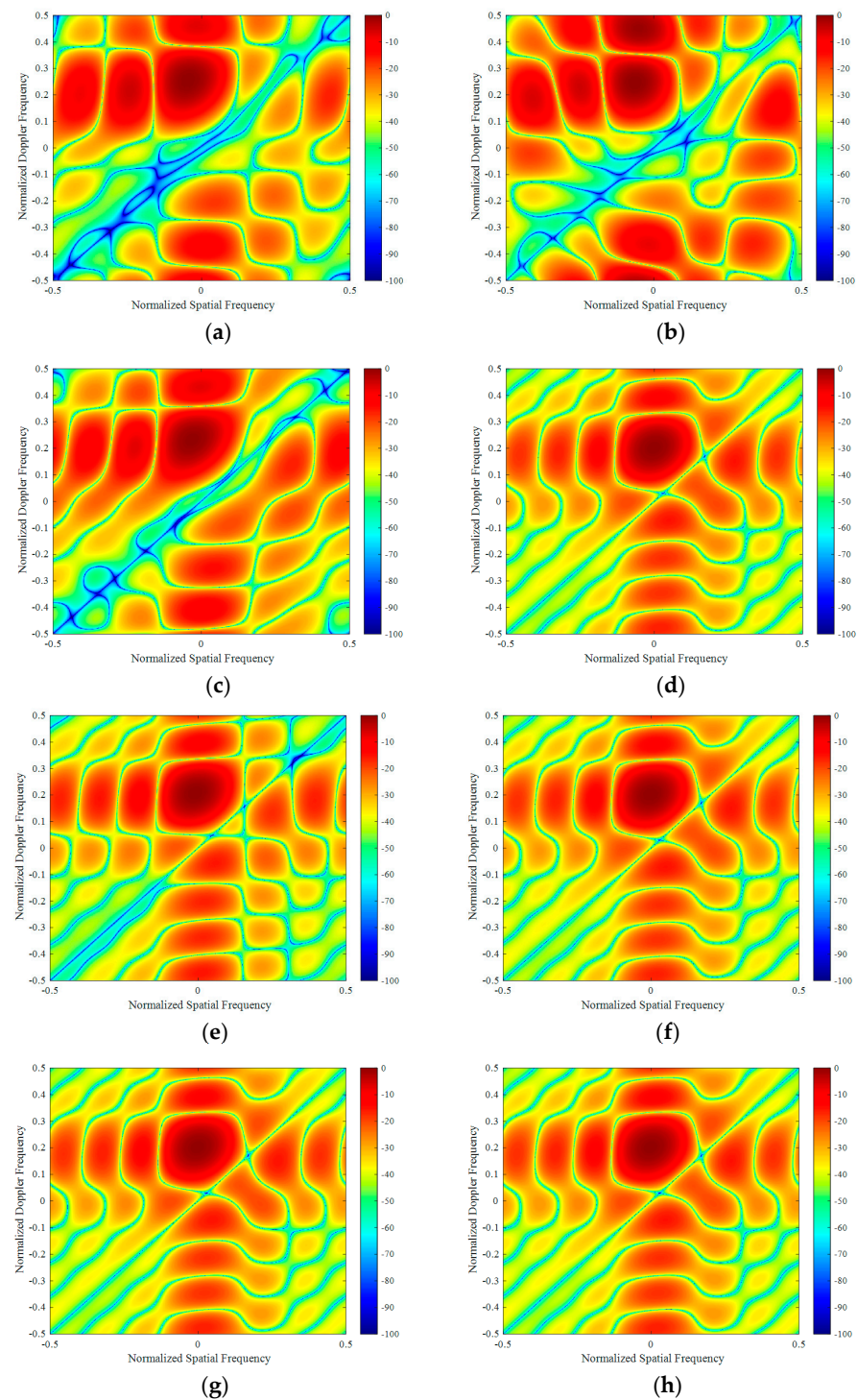
Figure 5. Spatial-temporal responses. (**a**) M-CVX; (**b**) M-OMP; (**c**) M-FOCUSS; (**d**) M-IAA; (**e**) M-SBL; (**f**) M-FCSBL; (**g**) M-SBL-IR$\ell_{2,1}$; (**h**) Proposed algorithm.

Note that because of the massive computational complexity of the M-CVX, it will not be discussed in the subsequent experiments. Then, Figure 6 illustrates the curves of the average SINR loss versus different a number of training samples. It should be noted that the average SINR loss is calculated by the mean of SINR loss for $f_d \in (-0.5, -0.05) \cup (0.05, 0.5)$ in this paper. As demonstrated in Figure 6, all algorithms have a certain degree of performance loss with one snapshot and increase with the number of training samples. ExCoV-IIR-MSBL apparently exhibits promising performance similar to the M-IAA and other SBL algorithms under small sample conditions. The proposed algorithm achieves a near-optimum performance with merely three training samples.
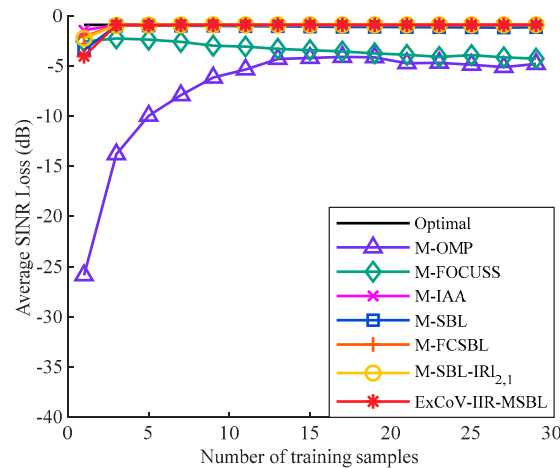
Figure 6. The average SINR loss versus the number of training samples in the ideal case.

Subsequently, we verify the target detection performance via the probability of detection (P_d) versus the target SNR curves, which are acquired by employing the cell-average constant false alarm rate (CA-CFAR) detector [37]. One hundred targets within the boresight are randomly added to the entire Range-Doppler plane, and the false alarm probability (P_{fa}) is fixed to 10^{-3}. Figure 7 shows that the proposed algorithm and other SBL based algorithms have noticeable improvements compared to M-OMP and M-FOCUSS. This result indicates that the proposed algorithm has superior target detection performance.
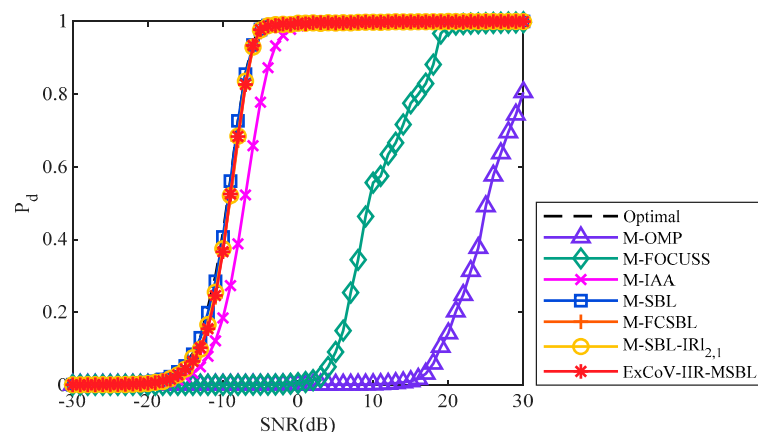

Figure 7. Probability of detection versus signal to noise ratio.

Besides, we also consider the SINR loss performance under the spatial mismatch case in the presence of angle-dependent gain-phase error. Specifically, the gain errors among the antennas are generated by a Gaussian distribution with a standard deviation of 0.03, and the phase errors are uniformly distributed within $[0°, 2°]$. Figure 8 shows the SINR loss versus normalized Doppler frequency. From Figure 8, the curves show that all the SR-STAP

algorithms suffer severe performance degradation due to the model mismatch between the dictionary and the actual clutter component. The M-IAA has the most serious performance loss, despite the fact that it can approach near-optimal as the SBL based algorithms in the ideal case.
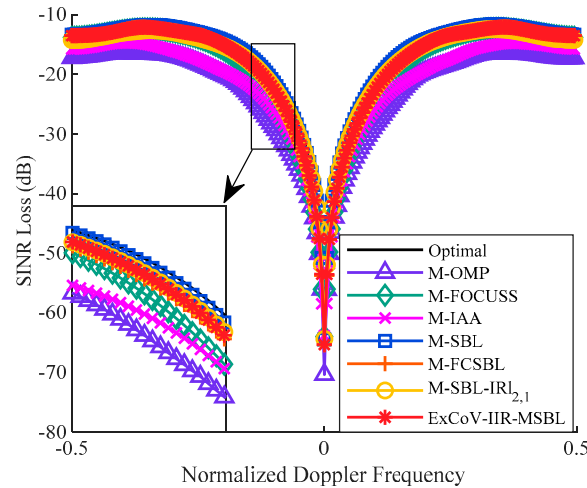

Figure 8. SINR loss comparison of different algorithms with gain-phase error.

Figure 9 shows the SINR loss versus the number of samples of above referred algorithms in the non-ideal case. As depicted in Figure 9, the proposed algorithm can yield a steady-state with only a few training samples as in the ideal case.
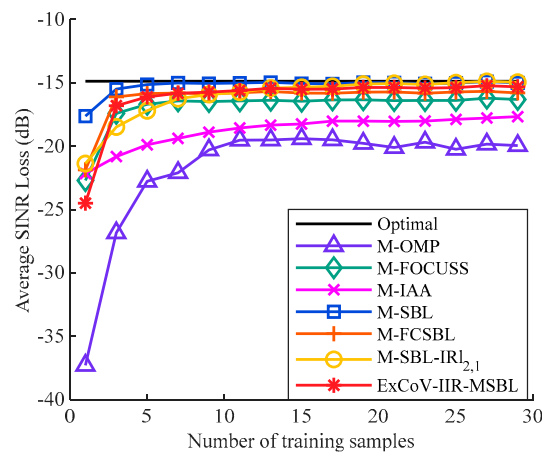

Figure 9. The average SINR loss versus the number of training samples with gain-phase error.

From the above experiments, it has been demonstrated that the proposed algorithm ensures a satisfactory clutter suppression and target detection performance while keeping a remarkable computational efficiency, thereby making it more suitable for implementation in modern large-scale radar systems.

7. Conclusions

In this work, to enhance the computational efficiency of the M-SBL based SR-STAP algorithm, we derive a novel algorithm called improved iterative reweighted sparse Bayesian learning based on expansion-compression variance-components (ExCoV-IIR-MSBL). Inspired by the intrinsic sparse prior to the clutter, an improved Bayesian probabilistic model with reduced parameter space is developed for SR-STAP. Furthermore, the GML criterion is utilized to partition the efficient and accurate parameter space. Then we extend the ExCoV method into the SR-STAP application to obtain the probabilistic model with

maximized GML objective function and enhance its convergence speed by a modified M-SBL-IR$\ell_{2,1}$ under the proposed Bayesian model. With numerical experiments, we show that the proposed algorithm outperforms other existing state-of-the-art algorithms with lower computational complexity.

Author Contributions: Conceptualization, W.C. and T.W.; investigation, D.W.; methodology, W.C. and D.W.; project administration, T.W.; software, W.C.; supervision, C.L.; visualization, W.C.; writing—original draft, W.C.; writing—review & editing, C.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Appendix A. Derivation of the FIM in (24)

In this Appendix, we develop the Fisher information matrix (FIM) for the unknown hyperparameters $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_{\Theta_\alpha} ; \boldsymbol{\gamma}_{\Theta_\beta} \right]$. The FIM is given by

$$\mathcal{I}(\boldsymbol{\psi}) = E_{p(\mathbf{X}|\boldsymbol{\psi})} \left\{ \left[\frac{\partial}{\partial \boldsymbol{\gamma}} \ln p(\mathbf{X}|\boldsymbol{\psi}) \right] \left[\frac{\partial}{\partial \boldsymbol{\gamma}} \ln p(\mathbf{X}|\boldsymbol{\psi}) \right]^T \right\} \tag{A1}$$

where $p(\mathbf{X}|\boldsymbol{\psi})$ is the joint probability density function of $\mathbf{X}$ conditioned on $\boldsymbol{\psi}$ and $E_{p(\mathbf{X}|\boldsymbol{\psi})}\{\cdot\}$ is the conditional expectation of $\mathbf{X}$ given $\boldsymbol{\psi}$. The derivative of $p(\mathbf{X}|\boldsymbol{\psi})$ with respect to the elements of $\boldsymbol{\gamma}$ is

$$\begin{aligned}
\frac{\partial}{\partial \gamma_k} \ln p(\mathbf{X}|\boldsymbol{\psi}) &= \frac{-L \ln |\mathbf{P}(\boldsymbol{\gamma})|}{\partial \gamma_k} - \frac{\partial \left(\sum\limits_{l=1}^{L} \mathbf{x}_l^H \mathbf{P}^{-1}(\boldsymbol{\gamma}) \mathbf{x}_l \right)}{\partial \gamma_k} \\
&= -L \cdot trace \left(\mathbf{P}^{-1}(\boldsymbol{\gamma}) \frac{\partial \mathbf{P}(\boldsymbol{\gamma})}{\partial \gamma_k} \right) - \sum\limits_{l=1}^{L} \mathbf{x}_l^H \frac{\partial \mathbf{P}^{-1}(\boldsymbol{\gamma})}{\partial \gamma_k} \mathbf{x}_l
\end{aligned} \tag{A2}$$

Put (A2) into (A1), and the $(k, j)th$ element of the FIM can be expressed as

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\psi})_{k,j} &= E_{p(\mathbf{X}|\boldsymbol{\psi})} \left\{ \left[\frac{\partial}{\partial \gamma_k} \ln p(\mathbf{X}|\boldsymbol{\psi}) \right] \left[\frac{\partial}{\partial \gamma_j} \ln p(\mathbf{X}|\boldsymbol{\psi}) \right] \right\} \\
&= E_{p(\mathbf{X}|\boldsymbol{\psi})} \left\{ \begin{array}{l} \left[-L \cdot trace \left(\mathbf{P}^{-1}(\boldsymbol{\gamma}) \frac{\partial \mathbf{P}(\boldsymbol{\gamma})}{\partial \gamma_k} \right) - \sum\limits_{l=1}^{L} \mathbf{x}_l^H \frac{\partial \mathbf{P}^{-1}(\boldsymbol{\gamma})}{\partial \gamma_k} \mathbf{x}_l \right] \\ \cdot \left[-L \cdot trace \left(\mathbf{P}^{-1}(\boldsymbol{\gamma}) \frac{\partial \mathbf{P}(\boldsymbol{\gamma})}{\partial \gamma_j} \right) - \sum\limits_{r=1}^{L} \mathbf{x}_r^H \frac{\partial \mathbf{P}^{-1}(\boldsymbol{\gamma})}{\partial \gamma_j} \mathbf{x}_r \right] \end{array} \right\} \\
&= 2L \cdot trace \left(\mathbf{P}(\boldsymbol{\gamma})^{-1} \frac{\partial \mathbf{P}(\boldsymbol{\gamma})}{\partial \gamma_k} \mathbf{P}^{-1}(\boldsymbol{\gamma}) \frac{\partial \mathbf{P}(\boldsymbol{\gamma})}{\partial \gamma_j} \right)
\end{aligned} \tag{A3}$$

Substitute (21) into (A3), the Equations (25)–(27) are then derived.

Appendix B. Parameter Update Step Derivation

The derivation of (30)–(32) are given in this Appendix. Since the first term $\ln|\mathbf{P}|$ in (14) is concave and non-decreasing with respect to $\boldsymbol{\gamma} \geq 0$. Then a quadratic upper bound auxiliary function of can be formed via the concave conjugate function [30].

$$\ln|\mathbf{P}| = \min_{z \geq 0} z^H \boldsymbol{\gamma} - h^*(z) \tag{A4}$$

where $h^*(z)$ denotes the concave conjugate of $\ln|\mathbf{P}|$, and z can be divided into $z = [z_\alpha; z_\beta]_{(K_\alpha + 1) \times 1}$.

Based on [19], the second term $\sum_{l=1}^{L} \mathbf{x}_l^H \mathbf{P}^{-1} \mathbf{x}_l$ can be equivalently represented as the following formulation.

$$\sum_{l=1}^{L} \mathbf{x}_l^H \mathbf{P}^{-1} \mathbf{x}_l = \min_{\mathbf{A}} \frac{1}{\sigma^2} \|\mathbf{X} - \mathbf{\Phi}\mathbf{A}\|_F^2 + \sum_{k \in \Theta_\alpha} \frac{\sum_{l=1}^{L} |\mathbf{A}_{kl}|^2}{\gamma_{\Theta_\alpha,k}} + \sum_{j \in \Theta_\beta} \frac{\sum_{l=1}^{L} |\mathbf{A}_{jl}|^2}{\gamma_{\Theta_\beta}} \tag{A5}$$

If we drop the minimization from (A4) and (A5), an upper bound on $\mathcal{L}(\boldsymbol{\gamma})$ in (14) can be obtained

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\gamma}) = \mathcal{L}(\boldsymbol{\gamma}_{\Theta_\alpha}, \gamma_{\Theta_\beta}) &= L \ln|\mathbf{P}| + \sum_{l=1}^{L} \mathbf{x}_l^H \mathbf{P}^{-1} \mathbf{x}_l \\
&\leq L \left(z_\alpha^H \boldsymbol{\gamma}_{\Theta_\alpha} + z_\beta^* \gamma_{\Theta_\beta} - h^*(z) \right) \\
&\quad + \sum_{k \in \Theta_\alpha} \frac{\sum_{l=1}^{L} |\mathbf{A}_{kl}|^2}{\gamma_{\Theta_\alpha,k}} + \sum_{j \in \Theta_\beta} \frac{\sum_{l=1}^{L} |\mathbf{A}_{jl}|^2}{\gamma_{\Theta_\beta}} + \frac{1}{\sigma^2} \|\mathbf{X} - \mathbf{\Phi}\mathbf{A}\|_F^2
\end{aligned} \tag{A6}$$

Following the basic principles of convex analysis, the optimal z is given by the gradient of $\ln|\mathbf{P}|$ with respect to $\boldsymbol{\gamma}$, i.e., $z = \nabla_{\boldsymbol{\gamma}} \ln|\mathbf{P}|$, then

$$z_i = \begin{cases} \mathbf{s}_i^H \mathbf{P}^{-1} \mathbf{s}_i & z_i \in z_\alpha \\ trace\left(\mathbf{P}^{-1} \mathbf{\Phi}_\beta \mathbf{\Phi}_\beta^H \right) & z_i = z_\beta \end{cases} \tag{A7}$$

Finally, for a fixed z, the Equations (30) and (31) are obtained by setting the derivative of the upper bound in (A6) with respect to $\boldsymbol{\gamma}$ to zero.

The noise power can be updated by utilizing the expectation-maximization (EM) Bayesian inference scheme.

$$\begin{aligned}
&E_{p(\mathbf{A}|\mathbf{X};\boldsymbol{\psi})} \left[\ln p(\mathbf{X}|\mathbf{A}, \sigma^2) \right] \\
&\propto E_{p(\mathbf{A}|\mathbf{X};\boldsymbol{\psi})} \left[-NML \ln \sigma^2 - \sigma^{-2} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\|_F^2 \right] \\
&= -NML \ln \sigma^2 - \sigma^{-2} E_{p(\mathbf{A}|\mathbf{X};\boldsymbol{\psi})} \left[\|\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\|_F^2 \right] \\
&= -NML \ln \sigma^2 - \sigma^{-2} \left(\|\mathbf{Y} - \mathbf{\Phi}\boldsymbol{\mu}\|_F^2 + L \cdot trace\left(\mathbf{\Phi}^H \mathbf{\Phi} \boldsymbol{\Sigma} \right) \right) \\
&= -NML \ln \sigma^2 - \sigma^{-2} \left(\|\mathbf{Y} - \mathbf{\Phi}\boldsymbol{\mu}\|_F^2 + L\sigma^2 \cdot trace\left(\mathbf{I} - \sigma^2 \mathbf{P}^{-1} \right) \right)
\end{aligned} \tag{A8}$$

Then Equation (32) is given by setting the derivative of (A8) with respect to σ^2 to zero.

References

1. Brennan, L.E.; Mallett, J.D.; Reed, I.S. Theory of Adaptive Radar. *IEEE Trans. Aerosp. Electron. Syst.* **1973**, *9*, 237–251. [CrossRef]
2. Klemm, R. *Principles of Space-Time Adaptive Processing*; The Institution of Electrical Engineers: London, UK, 2002.
3. Guerci, J.R. *Space-Time Adaptive Processing for Radar*; Artech House: Norwood, MA, USA, 2003.
4. Reed, I.S.; Mallett, J.D.; Brennan, L.E. Rapid Convergence Rate in Adaptive Arrays. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *10*, 853–863. [CrossRef]
5. Goldstein, J.S.; Reed, I.S. Reduced-rank adaptive filtering. *IEEE Trans. Signal Process.* **1997**, *45*, 492–496. [CrossRef]
6. Goldstein, J.S.; Reed, I.S. Theory of partially adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.* **1997**, *33*, 1309–1325. [CrossRef]
7. Haimovich, A.M. The eigencanceler: Adaptive radar by eigenanalysis methods. *IEEE Trans. Aerosp. Electron. Syst.* **1996**, *32*, 532–542. [CrossRef]
8. Wang, Y.L.; Liu, W.J.; Xie, W.C.; Zhao, Y.J. Reduced-rank space-time adaptive detection for airborne radar. *Sci. China Inf. Sci.* **2014**, *57*, 1–11. [CrossRef]
9. DiPietro, R. Extended factored space-time processing for airborne radar systems. *Signals Syst. Comput.* **1992**, *1*, 425–430.
10. Wang, H.; Cai, L. On adaptive spatial-temporal processing for airborne surveillance radar systems. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 660–670. [CrossRef]
11. Tong, Y.L.; Wang, T.; Wu, J.X. Improving EFA-STAP performance using persymmetric covariance matrix estimation. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 660–670. [CrossRef]

12. Shi, J.X.; Xie, L.; Cheng, Z.Y.; He, Z.S.; Zhang, W. Angle-Doppler channel selection method for reduced-dimension STAP based on sequential convex programming. *IEEE Commun. Lett.* **2021**, *25*, 3080–3084. [CrossRef]

13. Maria, S.; Fuchs, J. Application of the Global Matched Filter to Stap Data an Efficient Algorithmic Approach. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing, Toulouse, France, 14–19 May 2006.

14. Sun, K.; Zhang, H.; Li, G.; Meng, H.; Wang, X. A novel STAP algorithm using sparse recovery technique. In Proceedings of the IGARSS 2009, Cape Town, South Africa, 12–17 July 2009.

15. Sun, K.; Meng, H.; Wang, Y.; Wang, X. Direct data domain STAP using sparse representation of clutter spectrum. *Signal Process.* **2011**, *91*, 2222–2236. [CrossRef]

16. Han, S.D.; Fan, C.Y.; Huang, X.T. A Novel STAP Based on Spectrum-Aided Reduced-Dimension Clutter Sparse Recovery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 213–217. [CrossRef]

17. Duan, K.Q.; Wang, Z.T.; Xie, W.C.; Chen, H.; Wang, Y.L. Sparsity-based STAP algorithm with multiple measurement vectors via sparse Bayesian learning strategy for airborne radar. *IET Signal Process.* **2017**, *11*, 544–553. [CrossRef]

18. Wang, Z.T.; Xie, W.C.; Duan, K.Q.; Wang, Y.L. Clutter suppression algorithm based on fast converging sparse Bayesian learning for airborne radar. *Signal Process.* **2017**, *130*, 159–168. [CrossRef]

19. Liu, C.; Wang, T.; Zhang, S.; Ren, B. Clutter suppression based on iterative reweighted methods with multiple measurement vectors for airborne radar. *IET Radar Sonar Navig.* **2022**, *16*, 1446–1459. [CrossRef]

20. Yang, Z.C.; Li, X.; Wang, H.; Nie, L. Sparsity-based space-time adaptive processing using complex-valued Homotopy technique for airborne radar. *IET Signal Process.* **2014**, *8*, 552–564. [CrossRef]

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

22. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]

23. Yang, Z.C.; Li, X.; Wang, H.Q.; Jiang, W.D. On Clutter Sparsity Analysis in Space-Time Adaptive Processing Airborne Radar. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1214–1218. [CrossRef]

24. Zhang, Z.L.; Rao, B.D. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. *IEEE Trans. Signal Process.* **2013**, *61*, 2009–2015. [CrossRef]

25. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn.* **2001**, *1*, 211–244.

26. Qiu, K.; Dogandzic, A. Variance-component based sparse signal reconstruction and model selection. *IEEE Trans. Signal Process.* **2010**, *58*, 2935–2952. [CrossRef]

27. Ward, J. *Space-Time Adaptive Processing for Airborne Radar*; Lincoln Laboratory: Lexington, MA, USA, 1994.

28. Natarajan, B. Sparse approximate solutions to linear systems. *SIAM J. Comput.* **1995**, *24*, 227–234. [CrossRef]

29. Tipping, M.E.; Faul, A.C. Fast marginal likelihood maximization for sparse Bayesian models. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, USA, 3–6 January 2003; Volume 1, pp. 276–283.

30. Wipf, D.; Nagarajan, S. Iterative reweighted l1 and l2 methods for finding sparse solutions. *IEEE J. Sel. Top. Signal Process.* **2010**, *2*, 317–329. [CrossRef]

31. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1993.

32. Dogandzic, A.; Qiu, K. ExCoV: Expansion-compression variance-component based sparse-signal reconstruction from noisy measurements. In Proceedings of the 43rd Annual Conference Information Science and Systems, Baltimore, MA, USA, 18–20 March 2009; pp. 186–191.

33. Goodman, N.A.; Stiles, J.M. On clutter rank observed by arbitrary arrays. *IEEE Trans. Signal Process.* **2007**, *55*, 178–186. [CrossRef]

34. Tropp, J.A.; Gilbert, A.C.; Strauss, M.J. Algorithms for simultaneous sparse approximation, part I: Greedy pursuit. *Signal Process.* **2006**, *86*, 572–588. [CrossRef]

35. Tropp, J.A. Algorithms for simultaneous sparse approximation, part II: Convex relaxation. *Signal Process.* **2006**, *86*, 589–602. [CrossRef]

36. Cotter, S.F.; Rao, B.D.; Engan, K.; Kreutz-Delgado, K. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* **2005**, *53*, 2477–2488. [CrossRef]

37. Liu, W.J.; Liu, J.; Hao, C.P.; Gao, Y.C.; Wang, Y.L. Multichannel adaptive signal detection: Basic theory and literature review. *Sci. China Inf. Sci.* **2022**, *65*, 121301. [CrossRef]