remote sensing

MDPI

Article

A Novel Sparse Bayesian Space-Time Adaptive Processing Algorithm to Mitigate Off-Grid Effects

Cheng Liu, Tong Wang *, Kun Liu and Xinying Zhang

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

Abstract: Space-time adaptive processing (STAP) algorithms based on sparse recovery (SR) have been researched because of their low requirement of training snapshots. However, once some portion of clutter is not located on the grids, i.e., off-grid problems, the performances of most SR-STAP algorithms degrade significantly. Reducing the grid interval can mitigate off-grid effects, but brings strong column coherence of the dictionary, heavy computational load, and heavy storage load. A sparse Bayesian learning approach is proposed to mitigate the off-grid effects in the paper. The algorithm employs an efficient sequential addition and deletion of dictionary atoms to estimate the clutter subspace, which means that strong column coherence has no effect on the performance of the proposed algorithm. Besides, the proposed algorithm does not require much computational load and storage load. Off-grid effects can be mitigated with the proposed algorithm when the grid-interval is sufficiently small. The excellent performance of the novel algorithm is demonstrated on the simulated data.

Keywords: space-time adaptive processing; sparse recovery; off-grid; sparse Bayesian learning

Citation: Liu, C.; Wang, T.; Liu, K.; Zhang, X. A Novel Sparse Bayesian Space-Time Adaptive Processing Algorithm to Mitigate Off-Grid Effects. *Remote Sens.* **2022**, *14*, 3906. https://doi.org/10.3390/rs14163906

Academic Editors: Jingwei Xu, Keqing Duan, Weijian Liu and Xiongpeng He

Received: 14 July 2022
Accepted: 9 August 2022
Published: 11 August 2022

1. Introduction

Space-time adaptive processing (STAP) exhibits great potential in suppressing clutter and detecting slow-moving targets [1]. The Reed–Mallet–Brennan (RMB) rule [2] states that signal-to-clutter-plus-noise ratio (SCNR) loss can be within 3 dB only if the number of independent and identically distributed (i.i.d) training snapshots is at least twice the system degrees of freedom (DOFs). Unfortunately, due to array configurations and the complex clutter environment, sufficient training snapshots cannot be guaranteed, and therefore the performance of STAP degrades significantly.

Motivated by compressed sensing (CS) theory, STAP algorithms have been enriched over the past 15 years with the application of sparse recovery (SR) techniques. These algorithms are termed SR-STAP algorithms [3–9]. The clutter-plus-noise covariance matrix (CNCM) can be estimated with SR-STAP algorithms when the training snapshots are limited. Most of these algorithms are exploited to solve an ℓ_1-norm minimization problem instead of an ℓ_0-norm minimization problem [10]. However, large coefficients in the ℓ_1-norm are penalized more heavily, and their performances are greatly affected by the choice of regularization parameters. Compared with the above SR-STAP algorithms, sparse Bayesian learning (SBL) [11–21] has received much attention and has been applied to the STAP framework [22–24] because of its robustness and excellent performance. However, the performance of all SR-STAP algorithms relies on the match between clutter and the dictionary matrix.

The dictionary used in SR-STAP algorithms consists of the space-time steering vectors corresponding to all the discrete grids in the angular-Doppler plane. Based on the assumption that the clutter ridge lies exactly on the sampling grids, most SR-STAP algorithms aim to improve the accuracy of signal recovery. However, when this assumption does not hold,

some portion of clutter is not located on the grid, i.e., off-grid problems, which leads to inaccuracy of the CNCM estimated by SR-STAP algorithms and performance degradation [25]. Unfortunately, off-grid problems are unavoidable and common in practical applications.

Some algorithms have been proposed to overcome the off-grid problems in direction-of-arrival (DOA) estimation [26–29] but the off-grid problems in SR-STAP have not received attention. The first-order approximation method, which is common and effective in the off-grid DOA estimation, cannot be directly applied to off-grid SR-STAP because STAP is two-dimensional signal processing. In [25], to mitigate the off-grid effects, the dictionary is constructed by exploiting the knowledge of clutter ridge. Regrettably, precise environmental knowledge is hard to obtain and changes with time. In [30,31], the global atoms are selected from the global STAP dictionary and then the optimal atoms are searched from the local STAP dictionary. However, the convergence of these algorithms is not guaranteed.

It is obvious that an effective way to deal with the off-grid problems is to reduce the grid interval to obtain a denser grid set. However, traditional SR-STAP algorithms suffer from heavy computational load and storage load when utilizing the dense grid set. Moreover, if the grid-interval is small enough, the dictionary matrix cannot satisfy the restricted isometry property (RIP) condition [27], which also results in the inaccuracy of the CNCM and the performance degradation of SR-STAP algorithms.

In this paper, a novel SR-STAP algorithm is proposed to mitigate the effects of off-grid problems with sufficiently dense grids. The specific steps are as follows: Firstly, the angular-Doppler plane is discretized into sufficiently dense grids to construct an overcomplete dictionary; secondly, based on the marginal likelihood maximization criterion, we select the atoms that are in the clutter subspace and reserve these atoms in the new dictionary; thirdly, we estimate the CNCM and calculate the STAP weight vector.

Although sufficiently dense grids lead to strong column coherence of atoms in the dictionary, the core idea is to select the most accurate atoms to represent the clutter subspace based on the marginal likelihood maximization criterion, which indicates that the performance of the proposed cannot be affected by strong column coherence of atoms. Besides, due to the sparsity of clutter, the number of the reserved atoms is small, and the new dictionary is a low-dimensional matrix, which indicates that the proposed algorithm does not require a heavy computational load and storage load.

In conclusion, compared with the traditional SR-STAP algorithm, there are two main advantages of the proposed algorithm: (i) Compared with traditional SR-STAP algorithms, the run time and storage load of the proposed algorithm are both less in the absence and presence of off-grid problems; furthermore, (ii) the proposed algorithm can mitigate off-grid effects with sufficiently dense grids.

Notation: Scalar quantities, vectors, and matrices are denoted by italic typeface, boldface small letters, and boldface capital letters, respectively. The i-th entry of $\mathbf{x}$ is denoted by x_i. The i-th column and (i,j)-th element of $\mathbf{A}$ are denoted by $\mathbf{a}_i$ and $\mathbf{A}_{ij}$, respectively. The expectation is denoted by $\mathrm{E}(\cdot)$. $\mathrm{diag}(\cdot)$ denotes a diagonal matrix formed by a vector or diagonal elements of a matrix. The matrix transpose, conjugate transpose, and inverse are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{-1}$, respectively. $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and $\|\cdot\|_{2,0}$ denotes the ℓ_0-norm of the vector formed by the ℓ_2-norm of each row. $|\cdot|$ denotes the determinant. $\mathbf{I}$ denotes the identity matrix.

2. Signal Model

Without a loss of generality, an airborne pulsed-Doppler radar under consideration is equipped with a uniform linear array (ULA). N antenna elements are placed with a half-wavelength inner spacing. M pulses are transmitted at a constant pulse repetition frequency (PRF). The received signal $\mathbf{x}$ can be formulated as the superposition of clutter, thermal noise, and might targets. Ignoring the range ambiguity, the clutter $\mathbf{x}_{\mathrm{c}} \in \mathbb{C}^{MN\times 1}$

can be modelled as the superposition of echoes from all the clutter patches in the cell under test (CUT).

$$\mathbf{x}_{\mathrm{c}} = \sum_{i=1}^{N_{\mathrm{c}}} \alpha_i \boldsymbol{\upsilon}_{\mathrm{st}}(f_{\mathrm{d},i}, f_{\mathrm{s},i}) \tag{1}$$

$$\boldsymbol{\upsilon}_{\mathrm{st}}(f_{\mathrm{d},i}, f_{\mathrm{s},i}) = \boldsymbol{\upsilon}_{\mathrm{t}}(f_{\mathrm{d},i}) \otimes \boldsymbol{\upsilon}_{\mathrm{s}}(f_{\mathrm{s},i}) \tag{2}$$

$$\boldsymbol{\upsilon}_{\mathrm{t}}(f_{\mathrm{d},i}) = [1, \exp(j2\pi f_{\mathrm{d},i}), \ldots, \exp(j2\pi(M-1)f_{\mathrm{d},i})]^{\mathrm{T}} \tag{3}$$

$$\boldsymbol{\upsilon}_{\mathrm{s}}(f_{\mathrm{s},i}) = [1, \exp(j2\pi f_{\mathrm{s},i}), \ldots, \exp(j2\pi(N-1)f_{\mathrm{s},i})]^{\mathrm{T}} \tag{4}$$

where N_{c} is the number of clutter patches in the CUT; α_i, $f_{\mathrm{d},i}$, and $f_{\mathrm{s},i}$ are the complex amplitude, the normalized Doppler frequency, and the spatial frequency of the i-th clutter patch, respectively; the space-time steering vector $\boldsymbol{\upsilon}_{\mathrm{st}}$ can be obtained by the Kronecker product operation of the temporal steering vector $\boldsymbol{\upsilon}_{\mathrm{t}}$ and the spatial steering vector $\boldsymbol{\upsilon}_{\mathrm{s}}$; $\otimes$ denotes the Kronecker product operation.

In the sparsity-based STAP, we need to construct a dictionary matrix to represent the angular-Doppler plane. The angular-Doppler plane is discretized into $K = N_s N_d$ grids, where $N_s = \varphi_s N$ ($\varphi_s > 1$) and $N_d = \varphi_d M$ ($\varphi_d > 1$) are the number of normalized spatial frequency bins and Doppler frequency bins along the spatial frequency axes and the Doppler frequency axes, respectively. Each grid corresponds to a space-time steering vector, and the dictionary consists of all the space-time steering vectors. Assuming that clutter is located exactly on the grids, the received clutter plus noise snapshots $\mathbf{X} \in \mathbb{C}^{MN \times L}$ from L range cells in the multiple measurement vectors (MMV) case can be expressed as

$$\mathbf{X} = \mathbf{D}\mathbf{A} + \mathbf{N} \tag{5}$$

where $\mathbf{D} \in \mathbb{C}^{MN \times K}$, $\mathbf{A} \in \mathbb{C}^{K \times L}$, and $\mathbf{N} \in \mathbb{C}^{MN \times L}$ denote the dictionary matrix, the sparse coefficient matrix, and the noise matrix, respectively. $\mathbf{D}$ can be expressed as

$$\mathbf{D} = [\boldsymbol{\upsilon}_1, \boldsymbol{\upsilon}_2, \cdots, \boldsymbol{\upsilon}_K] \tag{6}$$

where

$$\boldsymbol{\upsilon}_k = \boldsymbol{\upsilon}_{\mathrm{t}}(f_{\mathrm{d},k}) \otimes \boldsymbol{\upsilon}_{\mathrm{s}}(f_{\mathrm{s},k}) \tag{7}$$

In (7), $f_{\mathrm{d},k}$ and $f_{\mathrm{s},k}$ are the normalized Doppler frequency and the spatial frequency of the k-th grid, respectively.

In the SR-STAP algorithms, the noise-contaminated $\mathbf{X}$ is required to be denoted with as few atoms as possible.

$$\mathbf{A} = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{A}\|_{2,0}, \quad \text{s.t.} \quad \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_{\mathrm{F}}^2 \leq \varepsilon \tag{8}$$

where ε is an error tolerance parameter related to the noise power.

3. Off-Grid Problems

In most SR-STAP algorithms, the grids are uniformly sampled in the angular-Doppler plane along the spatial frequency axes and the Doppler frequency axes, respectively. The performance of SR-STAP algorithms can be achieved on the assumption that the clutter ridge lies on the grids. Once some portion of the clutter ridge is not sampled by the grids, i.e., off-grid problems, the accuracy of the CNCM estimated by SR-STAP algorithms cannot be guaranteed, which leads to significant performance degradation. For example, (i) the slope of the clutter ridge is not equal to the ratio of N_s and N_d in the side-looking radar case and (ii) the clutter ridge of a non-side-looking radar is a non-linear curve.

Although a reduction in the grid interval can mitigate the effects of the off-grid problems, traditional SR-STAP algorithms suffer from heavy computational load and storage load. Moreover, if the grid- interval is small enough, the dictionary matrix cannot satisfy the restricted isometry property (RIP) condition. In the next section, a novel SR-STAP

algorithm is proposed to overcome the aforementioned problems when the grid- interval is reduced.

4. The Proposed Algorithm to Mitigate Off-Grid Effects

4.1. Construction of the Dictionary

The values of φ_s and φ_d can be set to an integer between 2 and 5 in the absence of off-grid problems. When off-grid problems occur, we need to reduce the grid interval by increasing the values of φ_s and φ_d in the proposed algorithm. In the presence of off-grid problems, the values of φ_s and φ_d can be set to an integer more than 10 or even higher.

Although precise environmental knowledge is hard to obtain, we still can know the approximate location of the clutter. To speed up the proposed algorithm, the grids that are definitely not in the approximate location of clutter can be removed from the dictionary at the beginning. However, the dictionary used in traditional SR-STAP needs to be over-completed. For a fair comparison, all SR-STAP algorithms exploit all grids in the whole angular-Doppler plane to construct the dictionary in the simulations.

4.2. Estimation of the Clutter Subspace

Assuming that the noise in (5) is complex Gaussian white noise, the likelihood function of $\mathbf{X}$ can be expressed as

$$p\left(\mathbf{X}\middle|\mathbf{A},\sigma^2\right)=\left(\pi\sigma^2\right)^{-MNL}\exp\left[-\sigma^{-2}\sum_{l=1}^{L}\|\mathbf{x}_l-\mathbf{D}\mathbf{a}_l\|^2\right] \tag{9}$$

where $\mathbf{A}=[\mathbf{a}_1,\mathbf{a}_2,\cdots,\mathbf{a}_L]$ and the noise variance σ^2 can be calculated with the knowledge of the radar system parameters.

Assuming that $\mathbf{a}_l$, $\forall l$ are submitted to the same zero-mean complex Gaussian prior distribution, the prior of $\mathbf{A}$ can be expressed as

$$p(\mathbf{A}|\mathbf{\Gamma})=\pi^{-KL}|\mathbf{\Gamma}|^{-L}\exp\left(-\sum_{l=1}^{L}\mathbf{a}_l^{\mathrm{H}}\mathbf{\Gamma}^{-1}\mathbf{a}_l\right) \tag{10}$$

where $\mathbf{\Gamma}=\mathrm{diag}(\boldsymbol{\gamma})$ represents the prior variance of $\mathbf{a}_l,\forall l$ and $\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\cdots,\gamma_K]^{\mathrm{T}}$ is a vector of hyper-parameters corresponding to all grids. Since $\mathbf{\Gamma}$ represents the prior variance of $\mathbf{a}_l,\forall l$, $\gamma_i\geq 0$.

Based on the Bayes rule, the posterior density of $\mathbf{A}$ can be expressed as

$$\begin{aligned}p\left(\mathbf{A}\middle|\mathbf{X},\mathbf{\Gamma},\sigma^2\right)&=\frac{p\left(\mathbf{X}\middle|\mathbf{A},\sigma^2\right)p(\mathbf{A}|\mathbf{\Gamma})}{\int p\left(\mathbf{X}\middle|\mathbf{A},\sigma^2\right)p(\mathbf{A}|\mathbf{\Gamma})d\mathbf{A}}\\&=\pi^{-KL}|\mathbf{\Sigma}|^{-L}\exp\left[-\sum_{l=1}^{L}(\mathbf{a}_l-\boldsymbol{\mu}_l)^{\mathrm{H}}\mathbf{\Sigma}^{-1}(\mathbf{a}_l-\boldsymbol{\mu}_l)\right]\\&=\prod_{l=1}^{L}\mathcal{CN}(\mathbf{a}_l|\boldsymbol{\mu}_l,\mathbf{\Sigma})\end{aligned} \tag{11}$$

where

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

$$\boldsymbol{\mu}_l=\sigma^{-2}\mathbf{\Sigma}\mathbf{D}^{\mathrm{H}}\mathbf{x}_l \tag{13}$$

The next step is to estimate $\boldsymbol{\gamma}$. If we adopt the expectation maximization (EM) approach to estimate $\boldsymbol{\gamma}$, then we can obtain the SBL algorithm with MMV (M-SBL) in [22]. Another

approach is that $\boldsymbol{\gamma}$ can be point estimated by maximizing the marginal likelihood function. The marginal likelihood function $p\left(\mathbf{X}\middle|\boldsymbol{\Gamma},\sigma^2\right)$ is expressed as

$$\begin{aligned} & p\left(\mathbf{X}\middle|\boldsymbol{\Gamma},\sigma^2\right) \\ & = \int p\left(\mathbf{X}\middle|\mathbf{A},\sigma^2\right) p(\mathbf{A}|\boldsymbol{\Gamma}) d\mathbf{A} \\ & = \pi^{-NML}\left|\mathbf{D}\boldsymbol{\Gamma}\mathbf{D}^{\mathrm{H}}+\sigma^2\mathbf{I}\right|^{-L} \exp\left[\sum_{l=1}^{L}\mathbf{x}_l^{\mathrm{H}}\left(\mathbf{D}\boldsymbol{\Gamma}\mathbf{D}^{\mathrm{H}}+\sigma^2\mathbf{I}\right)^{-1}\mathbf{x}_l\right] \end{aligned} \tag{14}$$

Define

$$\mathbf{C} \triangleq \mathbf{D}\boldsymbol{\Gamma}\mathbf{D}^{\mathrm{H}}+\sigma^2\mathbf{I} \tag{15}$$

and $\boldsymbol{\gamma}$ can be point estimated by

$$\begin{aligned} \boldsymbol{\gamma} & = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}} \ln p\left(\mathbf{X}\middle|\boldsymbol{\Gamma},\sigma^2\right) \\ & = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\left(-L\ln|\mathbf{C}|-\sum_{l=1}^{L}\mathbf{x}_l^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{x}_l\right) \end{aligned} \tag{16}$$

Define

$$\mathcal{L}(\boldsymbol{\Gamma}) \triangleq -L\ln|\mathbf{C}|-\sum_{l=1}^{L}\mathbf{x}_l^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{x}_l \tag{17}$$

and (16) can then be expressed as

$$\boldsymbol{\gamma} = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\mathcal{L}(\boldsymbol{\Gamma}) \tag{18}$$

An effective method to solve Equation (18) is to update a single hyper-parameter γ_j at a time. The update of γ_j makes the biggest contribution to the maximization of $\mathcal{L}(\boldsymbol{\Gamma})$. Next, we introduce how to select the serial number j.

$\mathbf{C}$ in (15) can be rewritten as

$$\begin{aligned} \mathbf{C} & = \sigma^2\mathbf{I}_{MN}+\sum_{1\le k\le K, k\ne i}\gamma_k\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{H}}+\gamma_i\boldsymbol{\nu}_i\boldsymbol{\nu}_i^{\mathrm{H}} \\ & = \mathbf{C}_{-i}+\gamma_i\boldsymbol{\nu}_i\boldsymbol{\nu}_i^{\mathrm{H}} \end{aligned} \tag{19}$$

where $\mathbf{C}_{-i} = \sigma^2\mathbf{I}_{MN}+\sum_{1\le k\le K, k\ne i}\gamma_k\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{H}}$ contains all terms that are independent of γ_i.

Using the Woodbury Matrix Identity

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1}-\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\left(\gamma_i^{-1}+\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\right)^{-1}\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1} \tag{20}$$

$$|\mathbf{C}| = |\mathbf{C}_{-i}|\left|1+\gamma_i\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\right| \tag{21}$$

Equation (17) can then be expressed as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Gamma}) & = -\sum_{l=1}^{L}\left[\mathbf{x}_l^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{x}_l+\ln|\mathbf{C}_{-i}|\right] \\ & \quad +\sum_{l=1}^{L}\left[\begin{array}{c} \mathbf{x}_l^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\left(\gamma_i^{-1}+\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\right)^{-1}\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{x}_l \\ -\ln\left|1+\gamma_i\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\boldsymbol{\nu}_i\right| \end{array}\right] \\ & = \mathcal{L}_{-i}+\ell(\gamma_i) \end{aligned} \tag{22}$$

where

$$\mathcal{L}_{-i} = -\sum_{l=1}^{L}\left[\mathbf{x}_l^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{x}_l+\ln|\mathbf{C}_{-i}|\right] \tag{23}$$

$$\ell(\gamma_i) = \sum_{l=1}^{L}\left[\mathbf{x}_l^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{v}_i\left(\gamma_i^{-1}+\mathbf{v}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{v}_i\right)^{-1}\mathbf{v}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{x}_l - \ln\left|1+\gamma_i\mathbf{v}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{v}_i\right|\right] \tag{24}$$

Equation (22) has been divided into two parts: The part independent of γ_i is denoted as $\mathcal{L}_{-i}$ and the other part related to γ_i is denoted as $\ell(\gamma_i)$.

Define

$$\hat{s}_i \triangleq \mathbf{v}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{v}_i \quad , \quad \hat{q}_{i,l} \triangleq \mathbf{v}_i^{\mathrm{H}}\mathbf{C}_{-i}^{-1}\mathbf{x}_l \tag{25}$$

and $\ell(\gamma_i)$ can then be simplified to

$$\ell(\gamma_i) = \sum_{l=1}^{L}\left[\hat{q}_{i,l}^{\mathrm{H}}\left(\gamma_i^{-1}+s_i\right)^{-1}\hat{q}_{i,l} - \ln|1+\gamma_i s_i|\right] \tag{26}$$

Differentiate $\ell(\gamma_i)$ with respect to γ_i

$$\frac{\partial\ell(\gamma_i)}{\partial\gamma_i} = -\sum_{l=1}^{L}\left[\hat{q}_{i,l}^{\mathrm{H}}(1+\gamma_i s_i)^{-2}\hat{q}_{i,l} - s_i(1+\gamma_i s_i)^{-1}\right] = 0 \tag{27}$$

and we can obtain the optimal γ_i^* by

$$\gamma_i^* = \begin{cases} \frac{1}{Ls_i^2}\sum\limits_{l=1}^{L}\left(\hat{q}_{i,l}^{\mathrm{H}}\hat{q}_{i,l} - s_i\right) & if \sum\limits_{l=1}^{L}\left(\hat{q}_{i,l}^{\mathrm{H}}\hat{q}_{i,l} - s_i\right) > 0_i \\ 0 & if \sum\limits_{l=1}^{L}\left(\hat{q}_{i,l}^{\mathrm{H}}\hat{q}_{i,l} - s_i\right) \le 0 \end{cases} \tag{28}$$

Let $\boldsymbol{\gamma}^* = \left[\gamma_1^*, \gamma_2^*, \cdots, \gamma_K^*\right]^{\mathrm{T}}$. In the $(t+1)$-th iteration, we need to identify the update of which hyper-parameter makes the biggest contribution to the maximization of $\mathcal{L}(\boldsymbol{\Gamma})$. The serial number of the corresponding hyper-parameter is j. Compared with other hyper-parameters, the update of the hyper-parameter γ_j makes the biggest contribution to the maximization of $\mathcal{L}(\boldsymbol{\Gamma})$. Therefore, j can be selected by the following equation

$$j = \underset{i,1\le i\le K}{\operatorname{argmax}}\mathcal{L}(\boldsymbol{\Gamma}) = \underset{i,1\le i\le K}{\operatorname{argmax}}\ \mathcal{L}_{-i} + \ell(\gamma_i^*) \tag{29}$$

where, herein $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma})$ and $\boldsymbol{\gamma} = \left[\gamma_1^{(t)}, \cdots, \gamma_i^*, \cdots, \gamma_K^{(t)}\right]^{\mathrm{T}}$.

In (29), $\mathcal{L}_{-i}, \forall i$ need to be calculated in each iteration. To improve efficiency, we define

$$\Delta\mathcal{L}(i) \triangleq \mathcal{L}_{-i} + \ell(\gamma_i^*) - \mathcal{L}\left(\boldsymbol{\Gamma}^{(t)}\right) = \mathcal{L}_{-i} + \ell(\gamma_i^*) - \mathcal{L}_{-i} - \ell\left(\gamma_i^{(t)}\right) = \ell(\gamma_i^*) - \ell\left(\gamma_i^{(t)}\right) \tag{30}$$

Equation (29) can be simplified as the following equation since $\mathcal{L}\left(\boldsymbol{\Gamma}^{(t)}\right)$ is a constant in the $(t+1)$-th iteration.

$$j = \underset{i,1\le i\le K}{\operatorname{argmax}}\left(\boldsymbol{\Gamma}^{(t)}\right) + \Delta\mathcal{L}(i) = \underset{i,1\le i\le K}{\operatorname{argmax}}\Delta\mathcal{L}(i) = \underset{i,1\le i\le K}{\operatorname{argmax}}\ell(\gamma_i^*) - \ell\left(\gamma_i^{(t)}\right) \tag{31}$$

Compared with (29), it is more convenient to select j with (31) and $\ell\left(\gamma_i^{(t)}\right)$, $\forall i$ been calculated in the last iteration. If j has been identified, replace γ_j with γ_j^* while fixing $\{\gamma_i | 1 \le i \le K, i \ne j\}$.

$$\gamma_j^{(t+1)} = \gamma_j^* \tag{32}$$

$$\gamma_i^{(t+1)} = \gamma_i^{(t)},\ 1 \le i \le K, i \ne j \tag{33}$$

As mentioned in the introduction, we select the atoms that are in the clutter subspace and reserve these atoms in the new dictionary. Therefore, the non-zero values in $\boldsymbol{\gamma}$ are reserved in $\boldsymbol{\Omega}$ and the corresponding atoms are reserved in the new dictionary $\boldsymbol{\psi}$.

$$\boldsymbol{\Omega} \triangleq \operatorname{diag}\left(\gamma_{\omega_1}, \gamma_{\omega_2}, \cdots, \gamma_{\omega_J}\right) \tag{34}$$

$$\boldsymbol{\psi} \triangleq \left[\mathbf{v}_{\omega_1}, \mathbf{v}_{\omega_2}, \cdots, \mathbf{v}_{\omega_J}\right] \tag{35}$$

where J is the number of atoms in $\boldsymbol{\psi}$ and

$$\left\{\omega_1, \omega_2, \cdots, \omega_J\right\} = \left\{i | 1 \leq i \leq K,\ \gamma_i \neq 0\right\} \tag{36}$$

We initialize $\boldsymbol{\gamma}^{(0)} = 0$, namely, $\boldsymbol{\Omega}^{(0)} = \varnothing$ and $\boldsymbol{\psi}^{(0)} = \varnothing$ at the beginning of the proposed algorithm. With the knowledge of $\gamma_j^{(t)}$ and $\gamma_j^{(t+1)}$, an efficient sequential addition and deletion of dictionary atoms can be taken to estimate $\boldsymbol{\psi}^{(t+1)}$ and $\boldsymbol{\Omega}^{(t+1)}$. If $\gamma_j^{(t)} = 0$ and $\gamma_j^{(t+1)} > 0$, $\boldsymbol{\psi}^{(t+1)} = [\boldsymbol{\psi}^{(t)}, \mathbf{v}_j]$ and $\boldsymbol{\Omega}^{(t+1)} = \begin{bmatrix} \boldsymbol{\Omega}^{(t)} & \\ & \gamma_j^{(t+1)} \end{bmatrix}$; if $\gamma_j^{(t)} > 0$ and $\gamma_j^{(t+1)} > 0$, $\boldsymbol{\psi}^{(t+1)} = \boldsymbol{\psi}^{(t)}$ and replace $\gamma_j^{(t)}$ with $\gamma_j^{(t+1)}$ in $\boldsymbol{\Omega}^{(t)}$; if $\gamma_j^{(t)} > 0$ and $\gamma_j^{(t+1)} = 0$, delete $\mathbf{v}_j$ from $\boldsymbol{\psi}^{(t)}$ and delete $\gamma_j^{(t)}$ from $\boldsymbol{\Omega}^{(t)}$; if $\gamma_j^{(t)} = 0$ and $\gamma_j^{(t+1)} = 0$, stop iteration because $\mathcal{L}(\boldsymbol{\Gamma})$ has already converged.

4.3. Fast Computation of $\left\{\hat{s}_i, \hat{q}_{i,l}\right\}$

Calculating the matrix inversion of $\mathbf{C}_{-i},\ \forall i$ brings a heavy computational load when $\left\{\hat{s}_i, \hat{q}_{i,l}\right\}$ are updated. Fast computation of $\left\{\hat{s}_i, \hat{q}_{i,l}\right\}$ is introduced as follows.

Define

$$s_i \triangleq \mathbf{v}_i^{\mathrm{H}} \mathbf{C}^{-1} \mathbf{v}_i\ ,\ q_{i,l} \triangleq \mathbf{v}_i^{\mathrm{H}} \mathbf{C}^{-1} \mathbf{x}_l \tag{37}$$

It is computationally efficient to calculate $\left\{\hat{s}_i, \hat{q}_{i,l}\right\}$ with $\left\{s_i, q_{i,l}\right\}$ because it is more convenient to calculate the matrix inversion of only $\mathbf{C}$ than that of all $\mathbf{C}_{-i},\ \forall i$.

With (20), we can obtain

$$\begin{aligned} s_i &= \mathbf{v}_i^{\mathrm{H}} \mathbf{C}^{-1} \mathbf{v}_i \\ &= \mathbf{v}_i^{\mathrm{H}} \left(\mathbf{C}_{-i}^{-1} - \mathbf{C}_{-i}^{-1} \mathbf{v}_i \left(\gamma_i^{-1} + \mathbf{v}_i^{\mathrm{H}} \mathbf{C}_{-i}^{-1} \mathbf{v}_i \right)^{-1} \mathbf{v}_i^{\mathrm{H}} \mathbf{C}_{-i}^{-1} \right) \mathbf{v}_i \\ &= \hat{s}_i - \hat{s}_i \left(\gamma_i^{-1} + \hat{s}_i \right)^{-1} \hat{s}_i \\ &= \frac{\hat{s}_i}{1 + \gamma_i \hat{s}_i} \end{aligned} \tag{38}$$

Thus, $\hat{s}_i$ can be calculated with s_i.

$$\hat{s}_i = \frac{s_i}{1 - \gamma_i s_i} \tag{39}$$

Similarly, $\hat{q}_{i,l}$ can be also calculated with $q_{i,l}$.

$$\hat{q}_{i,l} = (1 + \gamma_i \hat{s}_i) q_{i,l} = \frac{q_{i,l}}{1 - \gamma_i s_i} \tag{40}$$

To further improve the computational efficiency, we introduce the following approach to reduce computational complexities of $\left\{s_i, q_{i,l}\right\}$. With $\{\boldsymbol{\Omega}, \boldsymbol{\psi}\}$, (15) can be also expressed as

$$\mathbf{C} = \sigma^2 \mathbf{I}_{MN} + \boldsymbol{\psi} \Omega \boldsymbol{\psi}^{\mathrm{H}} \tag{41}$$

Define $\alpha \triangleq \sigma^{-2}$, and calculate $\{s_i, q_{i,l}\}$ with matrix inversion lemmas.

$$s_i = \alpha \boldsymbol{\nu}_i^{\mathrm{H}} \boldsymbol{\nu}_i - \alpha^2 \boldsymbol{\nu}_i^{\mathrm{H}} \boldsymbol{\psi} \boldsymbol{\Sigma} \boldsymbol{\psi}^{\mathrm{H}} \boldsymbol{\nu}_i \tag{42}$$

$$q_{i,l} = \alpha \boldsymbol{\nu}_i^{\mathrm{H}} \mathbf{x}_l - \alpha^2 \boldsymbol{\nu}_i^{\mathrm{H}} \boldsymbol{\psi} \boldsymbol{\Sigma} \boldsymbol{\psi}^{\mathrm{H}} \mathbf{x}_l \tag{43}$$

where $\boldsymbol{\Sigma}$ herein represents the covariance of $\boldsymbol{\chi}_l, \forall l$. $\boldsymbol{\chi}_l, \forall l$ is defined as the elements in $\mathbf{a}_l, \forall l$ whose corresponding hyper-parameters are non-zeros in the current iteration.

$$\boldsymbol{\Sigma} = \left(\sigma^{-2} \boldsymbol{\psi}^{\mathrm{H}} \boldsymbol{\psi} + \boldsymbol{\Omega}^{-1}\right)^{-1} \tag{44}$$

The mean of $\boldsymbol{\chi}_l, \forall l$ is expressed as

$$\boldsymbol{\mu}_l = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\psi}^{\mathrm{H}} \mathbf{x}_l, \ \forall l \tag{45}$$

With (42)~(45), we utilize $\{s_i, q_{i,l}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_l\}^{(t)}$ calculated in the t-th iteration to update $\{s_i, q_{i,l}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_l\}^{(t+1)}$ in the $(t+1)$-th iteration, which improves the computational efficiency. The updated formulas are listed in Appendix A.

4.4. Calculation of the STAP Filter Weight Vector

The CNCM $\mathbf{R}$ can be estimated by

$$\mathbf{R} = \frac{1}{L} \sum_{l=1}^{L} \sum_{i=1}^{J} \|\mu_{i,l}\|^2 \boldsymbol{\upsilon}_{\omega_i} \boldsymbol{\upsilon}_{\omega_i}^{\mathrm{H}} + \beta \sigma^2 \mathbf{I} \tag{46}$$

where $\beta \geq 1$ is a load factor.

The optimal STAP weight vector can be given by

$$\mathbf{w} = \frac{\mathbf{R}^{-1} \mathbf{s}_{\mathrm{t}}}{\mathbf{s}_{\mathrm{t}}^{\mathrm{H}} \mathbf{R}^{-1} \mathbf{s}_{\mathrm{t}}} \tag{47}$$

where $\mathbf{s}_{\mathrm{t}}$ is the steering vector of the target.

The proposed algorithm is shown in Algorithm 1. To reduce storage load, we do not store $\mathbf{D}$. When traversing the i-th grid, we use Formula (7) to generate the corresponding space-time steering vector.

Algorithm 1. Pseudocode for the proposed algorithm.

Step 1: Input: the data X, σ^2.
Step 2: Initialize: $\boldsymbol{\gamma}^{(0)} = 0$, $\mathbf{C} = \mathbf{C}_{-i} = \sigma^2 \mathbf{I}_{MN}$, and $\boldsymbol{\Omega}^{(0)} = \varnothing$, $\boldsymbol{\psi}^{(0)} = \varnothing$.
Step 3: While not converged do
  Obtain all $\gamma_i^*, \forall i$ by (28), and exploit (31) to find j-th hyper-parameter which needs to be updated in the current iteration.
  If $\gamma_j^{(t+1)} > 0$ and $\gamma_j^{(t)} = 0$,
  $\boldsymbol{\psi}^{(t+1)} = [\boldsymbol{\psi}^{(t)}, \boldsymbol{\nu}_j]$, and $\boldsymbol{\Omega}^{(t+1)} = \begin{bmatrix} \boldsymbol{\Omega}^{(t)} & \\ & \gamma_j^{(t+1)} \end{bmatrix}$.
  If $\gamma_j^{(t+1)} > 0$ and $\gamma_j^{(t)} > 0$,
  $\boldsymbol{\psi}^{(t+1)} = \boldsymbol{\psi}^{(t)}$, and replace $\gamma_j^{(t)}$ with $\gamma_j^{(t+1)}$.
  If $\gamma_j^{(t+1)} > 0$ and $\gamma_j^{(t)} > 0$,
  delete $\boldsymbol{\nu}_j$ from $\boldsymbol{\psi}^{(t)}$, and delete $\gamma_j^{(t)}$ from $\boldsymbol{\Omega}^{(t)}$.
  end
  Update $\{\boldsymbol{\Sigma}, \boldsymbol{\mu}, s_i, q_{i,l}\}, \forall i$ referring to Appendix A.
 end while
Step 4: Estimate the CNCM $\mathbf{R}$ by (46)
Step 5: Compute the space-time adaptive weight $\mathbf{w}$ using (47).
Step 6: The output of the space-time filter is $\mathbf{w}^{\mathrm{H}}\mathbf{X}$.

5. Analysis of Complexity, Storage and Convergence

5.1. Complexity Analysis

To reflect the low computational complexity of the proposed algorithm, some MMV SR-STAP algorithms, including the multiple focal underdetermined system solver (M-FOCUSS) and the M-SBL, are utilized as a comparison with the proposed algorithm. The number of multiplications for a single iteration is utilized as a measurement of the computational complexity. The dimension of $\boldsymbol{\psi}$ is $MN \times J$ and J is not fixed in each iteration. Because of the sparsity of clutter, $J < MN$. Suppose that J is equal to twice the rank of clutter when comparing the computational complexity of different algorithms in the simulations. The computational complexities of different algorithms are summarized in Table 1. For convenience, we set $\varphi_{\mathrm{s}} = \varphi_{\mathrm{d}}$, and define

$$\varphi \triangleq \varphi_{\mathrm{s}} = \varphi_{\mathrm{d}} \tag{48}$$

Table 1. The computational complexity for a single iteration.

Algorithm	Computational Complexity
M-FOCUSS	$o\left(MNKL + (MN)^3 + 2(MN)^2K\right)$
M-SBL	$o\left(MNKL + (MN)^3 + 2(MN)^2K + MNK^2\right)$
the proposed algorithm	$o\left((MN + 2MNJ + J^2 + L)K\right)$

Figure 1 illustrates the function of the computational complexities to φ in the side-looking radar case. In the simulations, $M = N = 10$, $L = 6$ and $\beta = 4v/\lambda f_{\mathrm{PRF}} = 1$. From Figure 1, the computational complexity of the proposed algorithm is far less than that of the M-FOCUSS and the M-SBL algorithm.

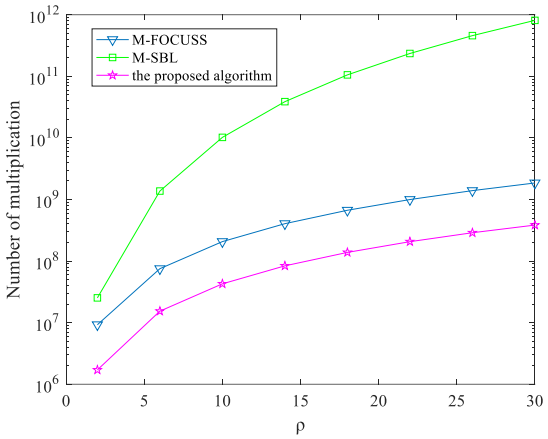

Figure 1. Computational complexity versus φ for a single iteration.

5.2. Storage Analysis

In order to reflect the low storage requirement of the proposed algorithm more intuitively, we compare the proposed algorithm with the M-SBL algorithms. We need to store $\mathbf{D} \in \mathbb{C}^{MN \times K}$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{K \times K}$ in the M-SBL algorithm while we store $\boldsymbol{\psi} \in \mathbb{C}^{MN \times J}$, $\boldsymbol{\Omega} \in \mathbb{R}^{J \times J}$ in the proposed algorithm. In the M-SBL algorithm, the dimension of $\boldsymbol{\Sigma}$ in (12) is $K \times K$ and the dimension of $\boldsymbol{\mu}_l$ in (13) is $K \times 1$. However, in the proposed algorithm, the dimension of $\boldsymbol{\Sigma}$ in (44) is $J \times J$ and the dimension of $\boldsymbol{\mu}_l$ in (45) is $J \times 1$. For example, when $M = N = 10$ and $\varphi_{\mathrm{s}} = \varphi_{\mathrm{d}} = 20$, $K = \varphi_{\mathrm{d}} M \varphi_{\mathrm{s}} N = 40{,}000$. In this case, the storage load in the M-SBL algorithm is heavy. Because of the sparsity of clutter, we can observe that $J < MN = 100$. The storage load in the proposed algorithm is far less than the M-SBL algorithm.

5.3. Convergence Analysis

According to [32], $\mathcal{L}(\mathbf{\Gamma})$ has an upper bound. We can conclude from (27) and (28) that $\ell\left(\gamma_j^{(t+1)}\right) = \ell\left(\gamma_j^{*}\right) \geq \ell\left(\gamma_j^{(t)}\right)$ and $\mathcal{L}\left(\mathbf{\Gamma}^{(t+1)}\right) \geq \mathcal{L}\left(\mathbf{\Gamma}^{(t)}\right), \forall t$. $\mathcal{L}(\mathbf{\Gamma})$ has an upper bound while it is also a monotonically increasing function, which means the proposed algorithm converges.

6. Performance Assessment

In this section, we verify the performance of the proposed algorithm and the other SR-STAP algorithms with the simulated data. The parameters of the airborne radar system are listed in Table 2. The SR-STAP algorithms for comparison are the M-FOCUSS and the M-SBL. We utilize the improvement factor (IF) as the measurement of performance.

$$\mathrm{IF} = \frac{\left|\mathbf{w}^{\mathrm{H}}\mathbf{s}\right|^2}{\mathbf{w}^{\mathrm{H}}\mathbf{R}\mathbf{w}} \frac{\mathrm{tr}(\mathbf{R})}{\mathbf{s}^{\mathrm{H}}\mathbf{s}} \tag{49}$$

Table 2. Parameters of airborne radar system.

Parameters	Symbols	Value
Distance between elements	d	0.15 m
Wavelength	λ	0.3 m
Platform height	H	9000 m
Number of pulses	M	8
Number of channels	N	8
Pulse repetition frequency	f_{PRF}	2000 Hz
Range sampling frequency	f_s	2.5 MHz
Clutter to noise ratio	CNR	40 dB

6.1. Comparison of Clutter Spectrums Estimated by SR-STAP Algorithms

Clutter spectrums estimated by the proposed algorithm and other SR-STAP algorithms are compared in the absence and presence of off-grid problems. We consider three different cases: (i) A side-looking radar without off-grid problems (the platform velocity is 150 m/s and $\beta = 1$); (ii) a side-looking radar with off-grid problems (the platform velocity is 180 m/s and $\beta = 1.2$); (iii) a forward-looking radar (the platform velocity is 180 m/s). When off-grid problems occur, the effects of off-grid problems are mitigated by increasing the value of φ. We also set $\varphi = 15$ and $\varphi = 25$ in the M-FOCUSS and M-SBL algorithms. However, either their running time is far beyond our acceptable range, the computer crashes, or their results are wrong because of the RIP condition. Therefore, their results are not shown when $\varphi = 15$ and $\varphi = 25$.

(i) A side-looking radar without off-grid problems

We compare the high-resolution spectrum estimated by the M-FOCUSS, the M-SBL, and the proposed algorithm. In the absence of off-grid problems, the value of φ can be set to an integer between 2 and 5. In the experiment, $\varphi = 5$. From the below figures in Figure 2, the clutter spectrums estimated by the M-SBL and the proposed algorithms are close to the ideal clutter spectrum in terms of the power and location of the clutter, which indicates the exact clutter power can be estimated by the proposed algorithm.

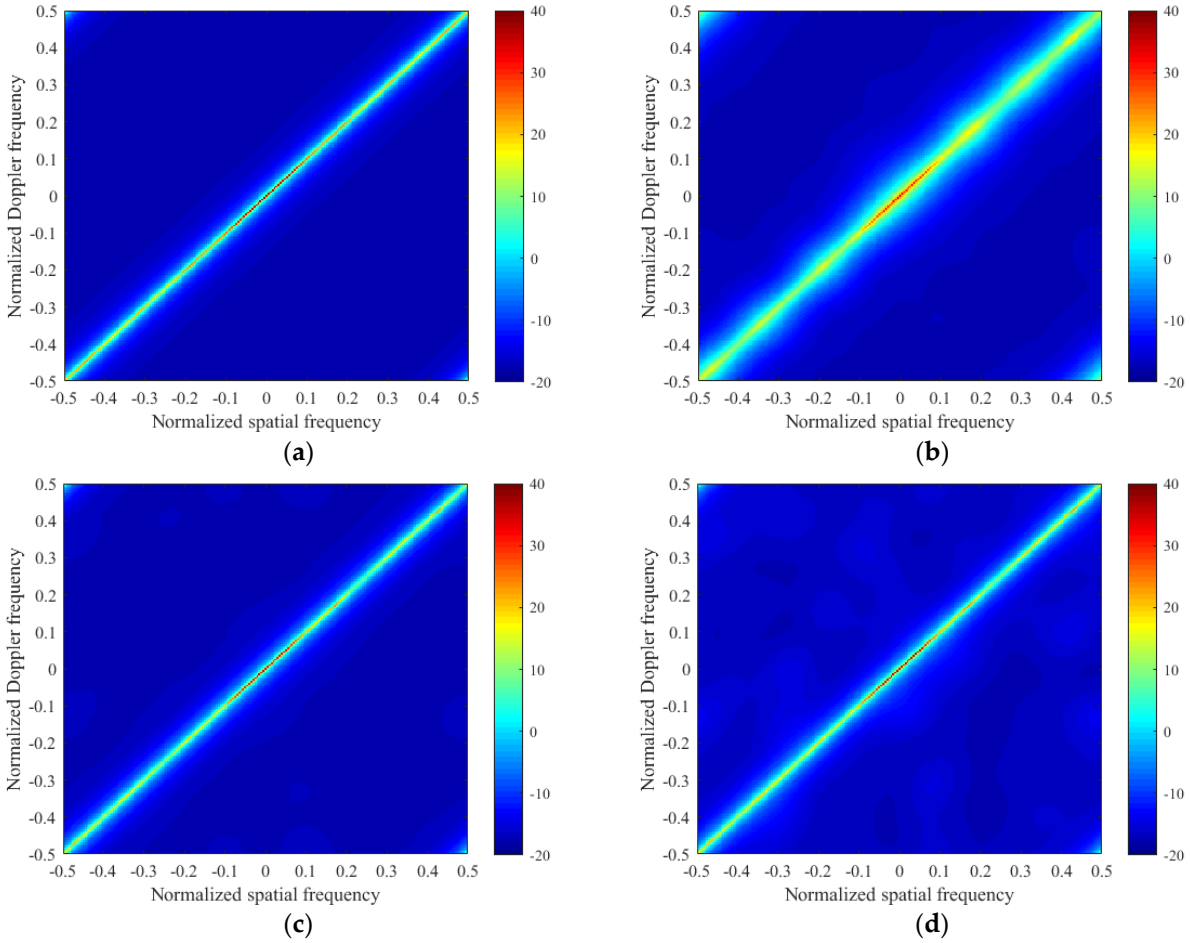

Figure 2. Comparison between the clutter angle-Doppler spectrum estimated by different algorithms for a side-looking radar without off-grid problems. (**a**) Ideal clutter spectrum; (**b**) M-FOCUSS, $\varphi = 5$; (**c**) M-SBL, $\varphi = 5$; (**d**) the proposed algorithm, $\varphi = 5$.

(ii) A side-looking radar with off-grid problems

Considering that the radar is a side-looking radar with off-grid problems, we set φ equal to 5, 15, and 25, respectively. When $\varphi = 15$ or $\varphi = 25$ in the M-FOCUSS and M-SBL algorithms, either their running time is far beyond our acceptable range, or the computer crashes, or their results are wrong because of the RIP condition. Therefore, their results are not shown when $\varphi = 15$ and $\varphi = 25$. From the below figures in Figure 3, the clutter spectrum estimated by the proposed algorithms is close to the ideal clutter spectrum in terms of the power and location of the clutter when $\varphi = 25$, which indicates that the exact clutter power can be estimated by the proposed algorithm with sufficiently dense grids. The experiment shows that the proposed algorithm can effectively mitigate the off-grid effects.

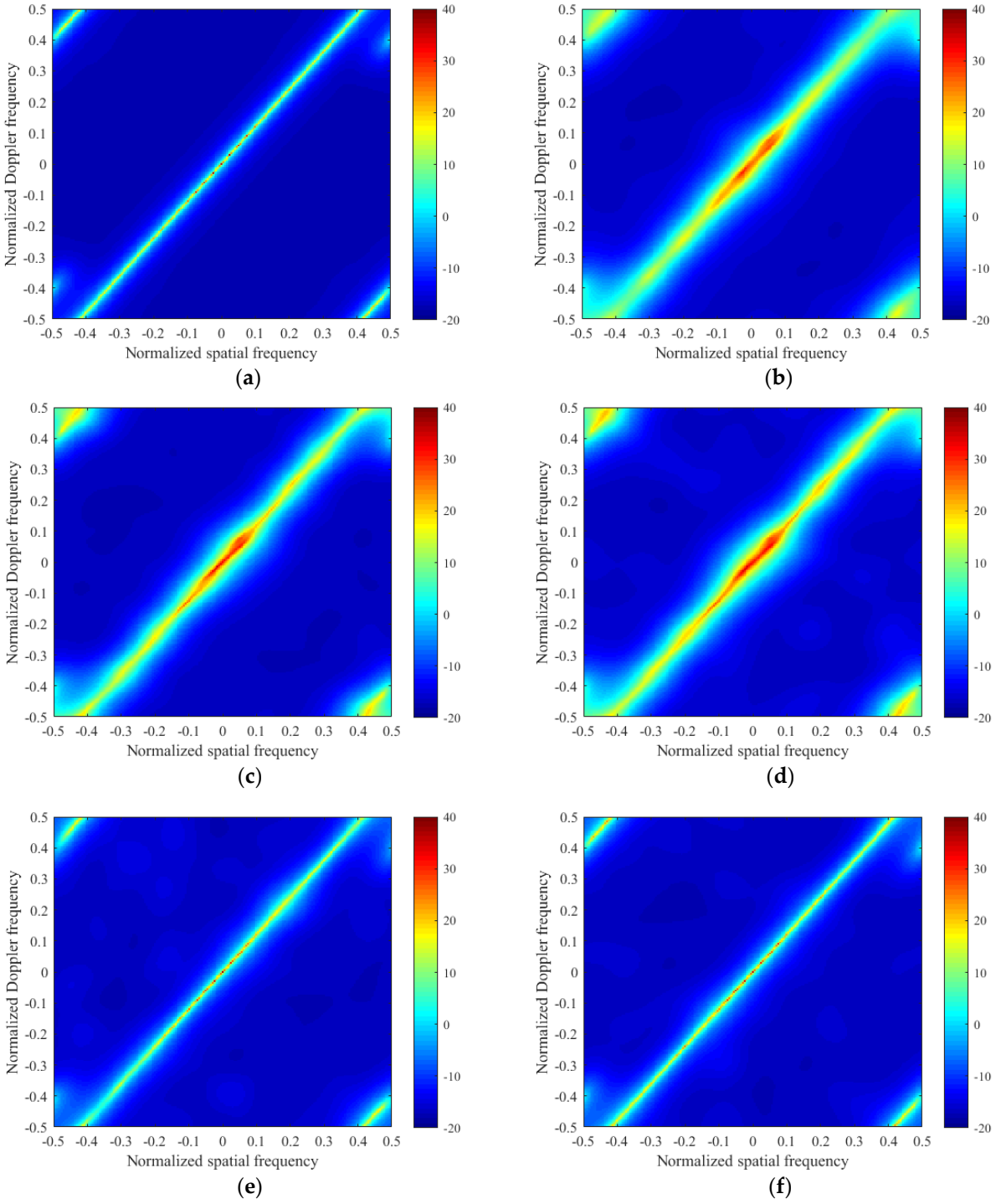

Figure 3. Comparison between the clutter angle-Doppler spectrum estimated by different algorithms for a side-looking radar with off-grid problems. (**a**) Ideal clutter spectrum; (**b**) M-FOCUSS, $\varphi = 5$; (**c**) M-SBL, $\varphi = 5$; (**d**) the proposed algorithm, $\varphi = 5$; (**e**) the proposed algorithm, $\varphi = 15$; (**f**) the proposed algorithm, $\varphi = 25$.

(iii) A forward-looking radar.

Considering that the radar is a forward-looking radar, we set φ equal to 5, 15, and 25, respectively. From the below figures in Figure 4, the clutter spectrum estimated by the proposed algorithms is close to the ideal clutter spectrum in terms of the power and location of the clutter when $\varphi = 25$, which indicates that the exact clutter power can be

estimated by the proposed algorithm with sufficiently dense grids. The experiment shows that the proposed algorithm can effectively mitigate the off-grid effects.

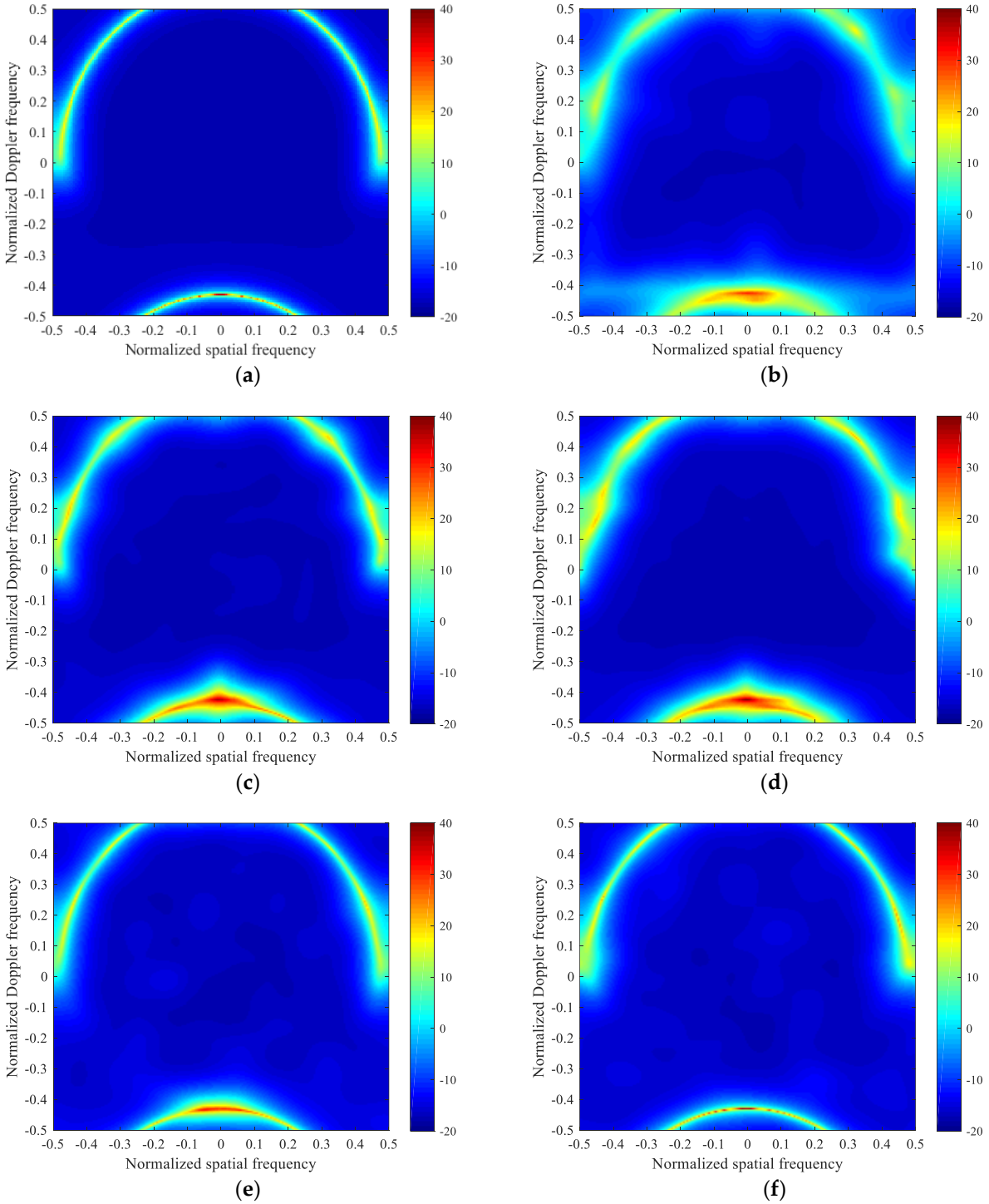

Figure 4. Comparison between the clutter angle-Doppler spectrum estimated by different algorithms for a forward-looking radar. (**a**) Ideal clutter spectrum; (**b**) M-FOCUSS, $\varphi = 5$; (**c**) M-SBL, $\varphi = 5$; (**d**) the proposed algorithm, $\varphi = 5$; (**e**) the proposed algorithm, $\varphi = 15$; (**f**) the proposed algorithm, $\varphi = 25$.

6.2. Comparison of IF Curves with SR-STAP Algorithms

In this experiment, we compare the clutter suppression performance of the proposed method with the M-FOCUSS and the M-SBL algorithms in the presence and absence of

off-grid problems. Furthermore, we also consider the ideal case and the non-ideal case. Amplitude Gaussian error (standard deviation 0.03) and phase random error (standard deviation 2°) are different in all directions in the non-ideal case.

(i) A side-looking radar without off-grid problems

In the absence of off-grid problems, the value of φ can be set equal to 5. From Figure 5, we note that the IF curves obtained by the M-SBL and the proposed algorithms are both close to the optimal IF curves obtained by the exact CNCM, which indicates that two algorithms can recover more exact clutter sources. The experiment shows that the proposed algorithm can also obtain good performance in the absence of off-grid problems.

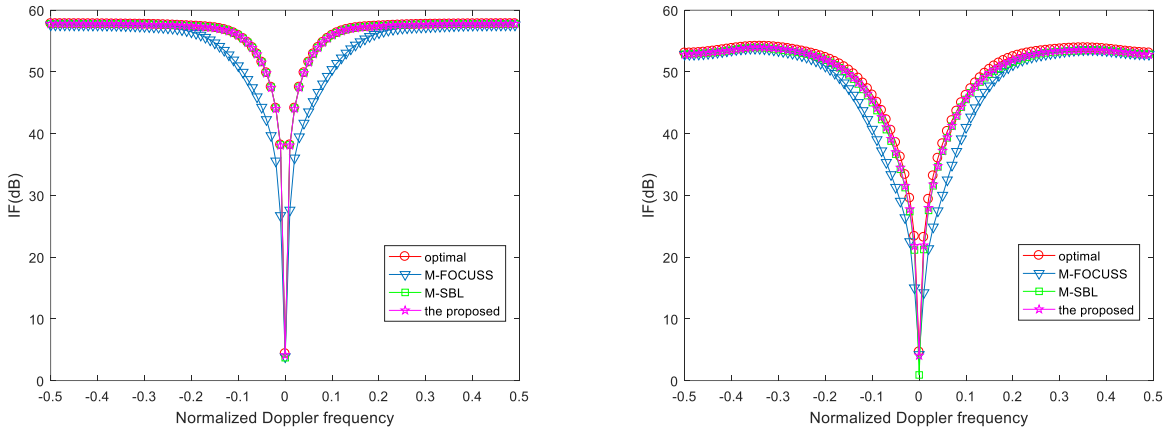

Figure 5. IF versus normalized Doppler frequency for a side-looking radar without off-grid problems. (**a**) Ideal case; (**b**) non-ideal case with array errors.

(ii) A side-looking radar with off-grid problems

Considering that the radar is a side-looking radar with off-grid problems, we set φ equal to 5, 15, and 25, respectively. From Figure 6, we find that increasing φ can effectively improve the clutter suppression performance in the presence of off-grid problems. When $\varphi = 25$, the IF curves obtained by the proposed algorithm are higher and narrower than the other IF curves. The experiment shows that the proposed algorithm can effectively mitigate the off-grid effects.

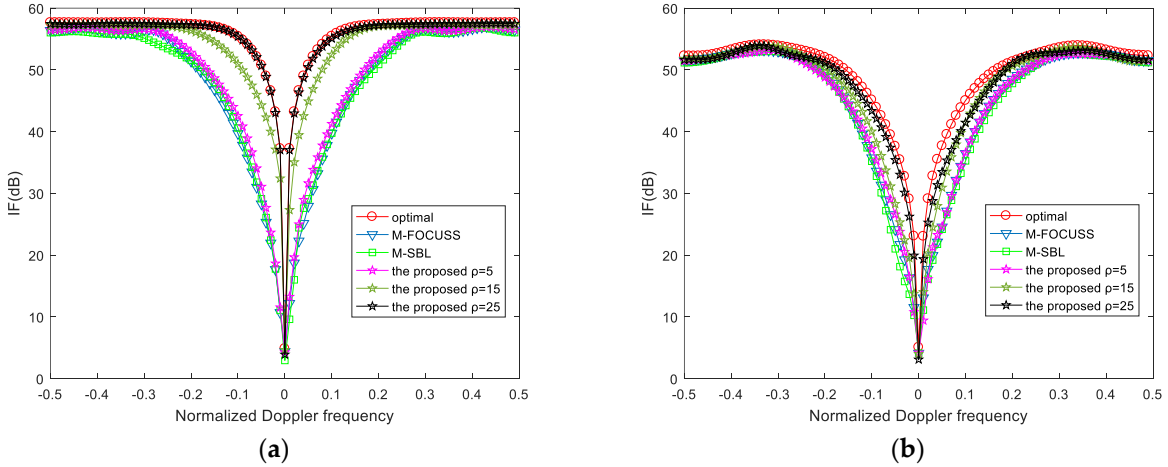

Figure 6. IF versus normalized Doppler frequency for a side-looking radar with off-grid problems. (**a**) Ideal case; (**b**) non-ideal case with array errors.

(iii) A forward-looking radar

Considering that the radar is a forward-looking radar, we set φ equal to 5, 15, and 25, respectively. From Figure 7, we find that increasing φ can effectively improve the clutter suppression performance in the presence of off-grid problems. When $\varphi = 25$, the IF curves obtained by the proposed algorithm are higher and narrower than the other IF curves. The experiment shows that the proposed algorithm can effectively mitigate the off-grid effects. Theoretically, as long as the grid points are dense enough, the off-grid problems can be solved with the proposed algorithm.

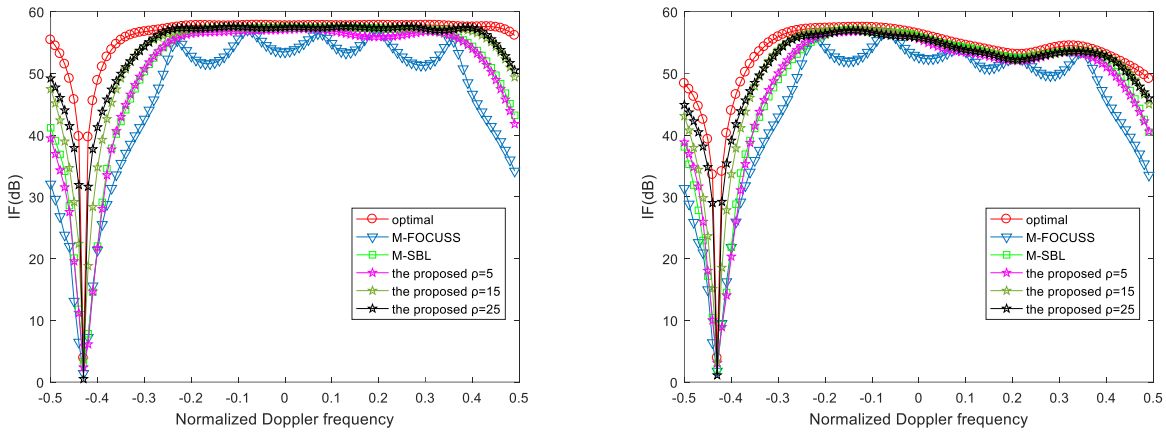

Figure 7. IF versus normalized Doppler frequency for a forward-looking radar. (**a**) Ideal case; (**b**) non-ideal case with array errors.

6.3. Comparison of Running Time with SR-STAP Algorithms

The symbol ∞ represents the running time is far beyond our acceptable range (1 hour) or beyond the capacity of the computer. We consider three different cases: (i) A side-looking radar without off-grid problems, (ii) a side-looking radar with off-grid problems, and (iii) a forward-looking radar. The running time of the algorithm is not only affected by the convergence performance of the algorithm itself, but also the computational complexity and the burden of the storage on the computer. The computer in the simulations is equipped with Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz 2.29GHz. The running time of the proposed algorithm in Tables 3–5 is far less than the traditional SR-STAP algorithms, especially when dense grids are exploited. One hundred Monte Carlo trials were performed to obtain the average running time.

(i) A side-looking radar without off-grid problems

Table 3. Running time of different algorithms for a side-looking radar without off-grid problems.

Algorithm	The Average Running Time (s)
M-FOCUSS	1.05
M-SBL	26.32
the proposed algorithm	0.88

(ii) A side-looking radar with off-grid problems

Table 4. Running time of different algorithms for a side-looking radar with off-grid problems.

Algorithm	The Average Running Time (s)		
	φ=5	φ=15	φ=25
M-FOCUSS	1.13	72.23	∞
M-SBL	29.79	∞	∞
the proposed algorithm	1.06	3.36	12.14

(iii) A forward-looking radar

Table 5. Running time of different algorithms for a forward-looking radar.

Algorithm	The Average Running Time (s)		
	φ=5	φ=15	φ=25
M-FOCUSS	1.02	64.14	∞
M-SBL	30.62	∞	∞
the proposed algorithm	1.05	4.03	12.94

7. Conclusions

A novel SR-STAP algorithm is proposed to mitigate off-grid effects. The proposed algorithm is based on the Bayes criterion and mitigates off-grid effects when the grid interval is sufficiently small. We pick out which atoms are in the subspace of clutter, which means the proposed algorithm does not need to satisfy the restricted isometry property (RIP) condition. Moreover, the complexity and storage requirement of the proposed algorithm is low, and its convergence can be promised.

Author Contributions: Conceptualization, T.W.; methodology, C.L.; validation, C.L., K.L. and X.Z.; writing—original draft preparation, C.L.; writing—review and editing, K.L. and X.Z.; visualization, C.L., K.L. and X.Z.; supervision, T.W.; project administration, T.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, grant number 2021YFA1000400.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics. You might choose to exclude this statement if the study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In order to make the following formulas more intuitive and concise, we remove (t) from quantities in the t-th iteration, and updated quantities in the $(t+1)$-th iteration are denoted by a tilde ($\sim$). As stated in the text, j represents the serial number of the hyper-parameter that needs to be updated in the $(t+1)$-th iteration. j also represents the position index of $\boldsymbol{\nu}_j$ in $\mathbf{D}$. When $\boldsymbol{\nu}_j$ is already in $\boldsymbol{\psi}$, g is defined as the position index of $\boldsymbol{\nu}_j$ in $\boldsymbol{\psi}$.

1. When $\widetilde{\gamma}_j > 0$ && $\gamma_j = 0$,

$$\Delta\mathcal{L}(j) = \ell(\widetilde{\gamma}_j) = \sum_{l=1}^{L}\left[\hat{q}_{j,l}^{\mathrm{H}}\left(\widetilde{\gamma}_j^{-1} + s_j\right)^{-1}\hat{q}_{j,l} - \ln\left|1 + \widetilde{\gamma}_j s_j\right|\right] \tag{A1}$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} + \alpha^2 z\boldsymbol{\Sigma}\boldsymbol{\psi}^{\mathrm{H}}\boldsymbol{\nu}_j\boldsymbol{\nu}_j^{\mathrm{H}}\boldsymbol{\psi}\boldsymbol{\Sigma} & -\alpha z\boldsymbol{\Sigma}\boldsymbol{\psi}^{\mathrm{H}}\boldsymbol{\nu}_j \\ -\alpha z\left(\boldsymbol{\Sigma}\boldsymbol{\psi}^{\mathrm{H}}\boldsymbol{\nu}_j\right)^{\mathrm{H}} & z \end{bmatrix} \tag{A2}$$

$$\tilde{\boldsymbol{\mu}}_l = \begin{bmatrix} \boldsymbol{\mu}_l - \alpha z q_{j,l}\boldsymbol{\Sigma}\boldsymbol{\psi}^{\mathrm{H}}\boldsymbol{\nu}_j \\ z q_{j,l} \end{bmatrix} \tag{A3}$$

$$\widetilde{s}_i = s_i - \alpha^2 z\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{e}_j\mathbf{e}_j^{\mathrm{H}}\boldsymbol{\nu}_i,\ \forall i \tag{A4}$$

$$\widetilde{q}_{i,l} = q_{i,l} - \alpha z q_{i,l}\boldsymbol{\nu}_i^{\mathrm{H}}\mathbf{e}_j,\ \forall i \tag{A5}$$

where $z = \left(\widetilde{\gamma}_j^{-1} + s_j\right)^{-1}$ and $\mathbf{e}_j = \boldsymbol{\nu}_j - \alpha\boldsymbol{\psi}\boldsymbol{\Sigma}\boldsymbol{\psi}^{\mathrm{H}}\boldsymbol{\nu}_j$.

2. When $\widetilde{\gamma}_j > 0$ && $\gamma_j > 0$,

$$\Delta\mathcal{L}(j) = \ell(\widetilde{\gamma}_j) - \ell(\gamma_j) = \sum_{l=1}^{L}\left[\begin{array}{l}\hat{q}_{j,l}^{\mathrm{H}}\left(\widetilde{\gamma}_j^{-1}+s_j\right)^{-1}\hat{q}_{j,l} - \hat{q}_{j,l}^{\mathrm{H}}\left(\gamma_j^{-1}+s_j\right)^{-1}\hat{q}_{j,l} \\ -\ln\left|1+\widetilde{\gamma}_j s_j\right| + \ln\left|1+\gamma_j s_j\right|\end{array}\right] \tag{A6}$$

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \kappa\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_g^{\mathrm{H}} \tag{A7}$$

$$\tilde{\boldsymbol{\mu}}_l = \boldsymbol{\mu} - \kappa\mu_{g,l}\boldsymbol{\Sigma}_g \tag{A8}$$

$$\widetilde{s}_i = s_i + \alpha^2\kappa\mathbf{v}_i^{\mathrm{H}}\boldsymbol{\psi}\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_g^{\mathrm{H}}\boldsymbol{\psi}^{\mathrm{H}}\mathbf{v}_i,\ \forall i \tag{A9}$$

$$\widetilde{q}_{i,l} = q_{i,l} + \alpha\kappa\mu_{g,l}\mathbf{v}_i^{\mathrm{H}}\boldsymbol{\psi}\boldsymbol{\Sigma}_g,\ \forall i \tag{A10}$$

where $\kappa = \left(-\frac{\gamma_j\widetilde{\gamma}_j}{\Delta\gamma_i} + \Sigma_{gg}\right)^{-1}$, $\Delta\gamma_j = \widetilde{\gamma}_j - \gamma_j$ and $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}(:,g)$.

3. When $\widetilde{\gamma}_j = 0$ && $\gamma_j > 0$,

$$\Delta\mathcal{L}(j) = -\ell(\gamma_j) = \sum_{l=1}^{L}\left[-\hat{q}_{j,l}^{\mathrm{H}}\left(\gamma_j^{-1}+s_j\right)^{-1}\hat{q}_{j,l} + \ln\left|1+\gamma_j s_j\right|\right] \tag{A11}$$

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \Sigma_{gg}^{-1}\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_g^{\mathrm{H}} \tag{A12}$$

$$\tilde{\boldsymbol{\mu}}_l = \boldsymbol{\mu} - \Sigma_{gg}^{-1}\mu_{g,l}\boldsymbol{\Sigma}_g \tag{A13}$$

$$\widetilde{s}_i = s_i + \alpha^2\Sigma_{gg}^{-1}\mathbf{v}_i^{\mathrm{H}}\boldsymbol{\psi}\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_g^{\mathrm{H}}\boldsymbol{\psi}^{\mathrm{H}}\mathbf{v}_i,\ \forall i \tag{A14}$$

$$\widetilde{q}_{i,l} = q_{i,l} + \alpha\Sigma_{gg}^{-1}\mu_{g,l}\mathbf{v}_i^{\mathrm{H}}\boldsymbol{\psi}\boldsymbol{\Sigma}_g,\ \forall i \tag{A15}$$

References

1. Ward, J. *Space-Time Adaptive Processing for Airborne Radar*; Technical Report; MIT Lincoln Laboratory: Lexington, KY, USA, 1998.
2. Reed, I.S.; Mallet, J.D.; Brennan, L.E. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *10*, 853–863. [CrossRef]
3. Donoho, D.L.; Elad, M.; Temlyakov, V.N. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **2006**, *52*, 6–18. [CrossRef]
4. Sun, K.; Zhang, H.; Li, G.; Meng, H.D.; Wang, X.Q. A novel STAP algorithm using sparse recovery technique. *IEEE Int. Geosci. Remote Sens. Symp.* **2009**, *1*, 3761–3764.
5. 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]
6. Sen, S. Low-rank matrix decomposition and spatial-temporal sparse recovery for STAP radar. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1510–1523. [CrossRef]
7. Yang, Z.; Wang, Z.; Liu, W.; de Lamare, R.C. Reduced-dimension space-time adaptive processing with sparse constraints on beam-Doppler selection. *Signal Process.* **2019**, *157*, 78–87. [CrossRef]
8. Zhang, W.; An, R.; He, N.; He, Z.; Li, H. Reduced dimension STAP based on sparse recovery in heterogeneous clutter environments. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *56*, 785–795. [CrossRef]
9. Liu, C.; Wang, T.; Zhang, S.G.; Ren, B. Clutter suppression based on iterative reweighted methods with multiple measurement vectors for airborne radar. *IET Radar Sonar Navig.* **2022**, *16*, 1–14. [CrossRef]
10. Candes, M.; Wakin, M.; Boyd, S. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **2008**, *5*, 877–905. [CrossRef]
11. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn.* **2001**, *1*, 211–244.
12. Wipf, D.P.; Rao, B.D. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153–2164. [CrossRef]
13. Wipf, D.P.; Rao, B.D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* **2007**, *55*, 3704–3716. [CrossRef]
14. 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.
15. Ji, S.H.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process.* **2008**, *56*, 2346–2356. [CrossRef]
16. Ji, S.H.; Dunson, D.; Carin, L. Multi-task compressive sensing. *IEEE Trans. Signal Process.* **2009**, *57*, 92–106. [CrossRef]
17. Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Bayesian compressive sensing using Laplace priors. *IEEE Trans. Image Process.* **2010**, *19*, 53–63. [CrossRef]

18. Wu, Q.S.; Zhang, Y.M.; Amin, M.G.; Himed, B. Complex multitask Bayesian compressive sensing. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014.
19. Serra, J.G.; Testa, M.; Katsaggelos, A.K. Bayesian K-SVD using fast variational inference. *IEEE Trans. Image Process.* **2017**, *26*, 3344–3359. [CrossRef]
20. Ma, Z.Q.; Dai, W.; Liu, Y.M.; Wang, X.Q. Group sparse Bayesian learning via exact and fast marginal likelihood maximization. *IEEE Trans. Signal Process.* **2017**, *65*, 2741–2753. [CrossRef]
21. Liu, C.; Wang, T.; Zhang, S.G.; Ren, B. A fast space-time adaptive processing algorithm based on sparse Bayesian learning for airborne radar. *Sensors* **2022**, *22*, 2664. [CrossRef]
22. 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]
23. Wang, Z.T.; Xie, W.C.; Duan, K.Q. Clutter suppression algorithm base on fast converging sparse Bayesian learning for airborne radar. *Signal Process.* **2017**, *130*, 159–168. [CrossRef]
24. Yang, X.P.; Sun, Y.Z.; Yang, J.; Long, T.; Sarkar, T.K. Discrete Interference suppression method based on robust sparse Bayesian learning for STAP. *IEEE Access* **2019**, *10*, 26740–26751. [CrossRef]
25. Duan, K.Q.; Liu, W.J.; Duan, G.Q.; Wang, Y.L. Off-grid effects mitigation exploiting knowledge of the clutter ridge for sparse recovery STAP. *IET Radar Sonar Navig.* **2018**, *12*, 557–564. [CrossRef]
26. You, K.T.; Guo, W.B.; Liu, Y.L.; Wang, W.B.; Sun, Z. Grid evolution: Joint dictionary learning and sparse Bayesian recovery for multiple off-grid targets localization. *IEEE Commun. Lett.* **2018**, *22*, 2068–2071. [CrossRef]
27. Dai, J.S.; Bao, X.; Xu, W.C.; Chang, C.Q. Root sparse Bayesian learning for off-grid DOA estimation. *IEEE Signal Process. Lett.* **2017**, *24*, 46–50. [CrossRef]
28. Fang, J.; Wang, F.Y.; Shen, Y.N.; Li, H.B.; Blum, R.S. Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach. *IEEE Trans. Signal Process.* **2016**, *64*, 4649–4662. [CrossRef]
29. Fang, J.; Shen, Y.N.; Li, H.B.; Li, S.Q. Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and sparse signal recovery. *IEEE Signal Process. Lett.* **2014**, *21*, 761–766.
30. Li, Z.H.; Zhang, Y.S.; He, X.Y.; Guo, Y.D. Low-complexity off-grid STAP algorithm based on local search clutter subspace estimation. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1862–1865. [CrossRef]
31. Yuan, H.D.; Xu, H.; Duan, K.Q.; Xie, W.C.; Liu, W.J.; Wang, Y.L. Sparse Bayesian learning-based space-time adaptive processing with off-grid self-calibration for airborne radar. *IEEE Access* **2018**, *6*, 47296–47307. [CrossRef]
32. Wipf, D.; Nagarajan, S. A new view of automatic relevance determination. In *Advances in Neural Information Processing Systems 20*; MIT Press: New York, NY, USA, 2008.