IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 26, 2022, accepted February 23, 2022, date of publication March 3, 2022, date of current version March 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3156638

Multi-Task Learning STAP via Spatial Smoothness and Group Sparsity Regularizations

LILONG QIN, BO TANG, HAI WANG, AND ZHONGRUI HUANG
College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China

Corresponding author: Lilong Qin (qinlilong13@nudt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801500 and Grant 62171450, and in part by the Anhui Provincial Natural Science Foundation under Grant 1908085QF252 and Grant 2108085J30.

ABSTRACT In practical applications, limited independent and identically distributed training snapshots brings a serious challenge in space-time adaptive processing (STAP), especially in the nonhomogeneous environments. Motivated by the significant spatial smoothness and sparsity commonality of weight vectors among related STAP tasks, we propose a novel STAP algorithm based on multi-task learning. In the proposed algorithm, the weight vectors corresponding to neighboring range bins of interest are kept consistent, and all weight vectors are constrained to share a common feature. Then, an alternating direction method of multipliers (ADMM) is used to solve the proposed algorithm, and the convergence of the algorithm is guaranteed. In addition, in case that the feature matrix is unknown or we want to learn a better feature matrix so that the associations among STAP tasks can be enhanced, we also provide an extension of the proposed algorithm to jointly optimize the feature matrix and weight matrix. Simulation results demonstrate the effectiveness of the proposed strategies.

INDEX TERMS ADMM, convex optimization, group-sparsity, multi-task learning, STAP.

I. INTRODUCTION

Space-time adaptive processing (STAP) can improve the performance of slow-moving target indication effectively for airborne/spaceborne radar systems in the presence of strong ground/sea clutter and interference (see, e.g., [1]–[3]). In conventional STAP algorithms, in order to ensure that the signal-to-clutter-plus-noise ratio (SCNR) loss is less than 3dB, the required number of independent and identically distributed (IID) training snapshots should be at least twice the number of system degrees of freedom (DoFs) [1]. Generally, this is on the order of several hundreds, which far exceeds the available training snapshots because of the environment heterogeneity [4].

Up to now, many strategies have been proposed to get rid of the limitations of conventional STAP algorithms. Reduced-rank (RR) STAP strategies can reduce the required number of training snapshots while sustaining the detection performance of conventional STAP algorithms [5], [6]. Nevertheless, the prior knowledge of the rank of clutter-plus-jamming subspace is required, which may be difficult to determine. Moreover, eigenvalue decomposition in RR-STAP strategies is computationally expensive. In order to reduce the number of snapshots and the computational complexity simultaneously, reduced-dimension (RD) STAP strategies have been presented [7]–[9]. Nonetheless, the nonadaptive construction of projection matrix, which depends on the selection of intuitive experience, degrades the detection performance [1].

Recently, sparsity-regularized STAP methods have been studied, and the theoretical analysis using the ℓ_1-norm regularization constraint has been established [10], [11]. To effectively solve the sparsity-regularized STAP problem, the ℓ_1-regularized recursive least-squares STAP algorithm [12], the ℓ_1-regularized least-mean-square STAP algorithm [13], and the Homotopy-STAP algorithm [14] have been proposed. Compared with conventional STAP algorithms, sparsity-regularized STAP algorithms exhibit a better performance [15].

However, the prior knowledge is not taken into consideration by the algorithms mentioned above. When the dimension of optimization problem is high, optimization algorithms which do not use the prior knowledge will perform poorly to estimate the parameters reliably. Knowledge-aided (KA) STAP incorporates a priori knowledge into the

The associate editor coordinating the review of this manuscript and approving it for publication was Guolong Cui.

estimation process to accelerate the convergence of the covariance matrix [16]–[19]. The acquisition of prior knowledge relies on the digital terrain elevation data, land cover land use data, synthetic aperture radar data or even hyperspectral imagery data of interest area, whereas these data are sometimes difficult to obtain or might lose their relevance with time [20].

Motivated by the significant spatial smoothness and sparsity commonality of weight vectors among related STAP tasks, in this paper, we propose two novel multi-task learning STAP algorithms to accelerate the convergence. Multi-task learning is especially beneficial when the number of training sample is small for each STAP task [21], [22]. Fig. 1 illustrates the difference between traditional single task learning (STL) STAP and multi-task learning (MTL) STAP. In STL-STAP, each STAP task is considered to be independent, and the weight vector is learnt independently. In MTL-STAP, multiple STAP tasks are learnt simultaneously. By utilizing task relevance, the number of training snapshots for each task is increased equivalently, and the detecting performance can be improved. Specifically, the contributions of this paper are:

(1) Spatial smoothness and sparsity commonality of weight vectors among multiple related STAP tasks are analyzed, and a multi-task learning STAP algorithm is proposed to provide a better performance with a small number of IID training snapshots;

(2) Alternating direction method of multipliers (ADMM) is used to solve the optimazation problem, and the convergence is guaranteed;

(3) In the case that the feature matrix is unknown or we want to learn a better feature matrix so that the associations among STAP tasks is enhanced, an extension of the proposed algorithm, i.e., multi-task feature learning STAP algorithm, is proposed to jointly optimize the feature matrix and weight matrix, and an equivalent convex optimization problem is provided.

The rest of paper is organized as follows. Section II briefly recalls the system model of sparsity-regularized STAP, introduces the spatial smoothness and sparsity commonality of weight vectors among related STAP tasks. Section III develops a multi-task learning STAP algorithm, and ADMM is used to tackle the optimization problem. An extension of the proposed algorithm to jointly optimize the feature matrix and weight matrix is provided in Section IV. Section V provides some numerical simulations to demonstrate the effectiveness of the proposed algorithms. Finally, we draw the conclusions in Section VI.

Notations: Throughout this paper, variables, vectors, and matrices are represented by lowercase letters, lowercase bold letters, and uppercase bold letters, respectively. Transpose, complex conjugation, and conjugate transpose are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, and $(\cdot)^{\mathrm{H}}$, respectively. The symbol $\otimes$ denotes the Kronecker product. The symbol $\|\cdot\|_n$ denotes the ℓ_n-norm operator, and the $\ell_{r,p}$-norm is defined as

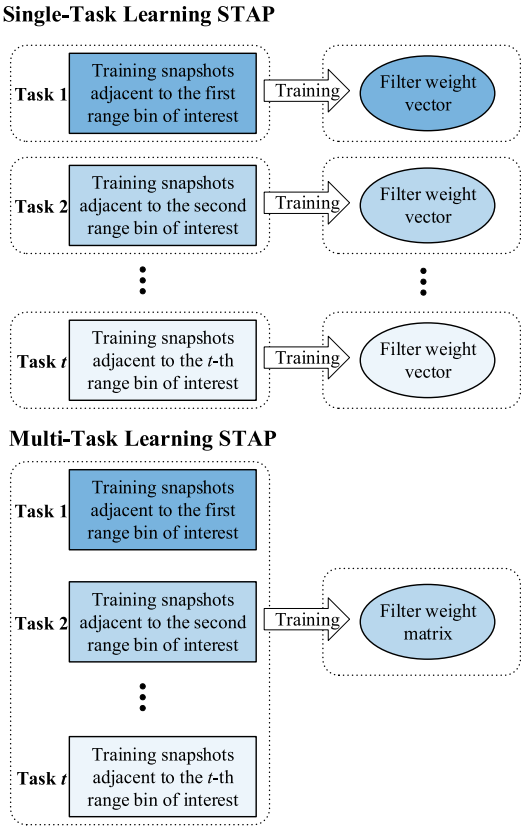


FIGURE 1. Illustration of STL-STAP and MTL-STAP.

$\|\mathbf{X}\|_{r,p} = \left(\sum_{d=1}^{D} \|\mathbf{x}_d\|_r^p\right)^{\frac{1}{p}}$, where $\mathbf{x}_d$ denotes d-th row in $\mathbf{X}$. Hence, the $\ell_{2,1}$-norm enforces row-wise sparsity. $E(\mathbf{x})$ denotes the expected value of $\mathbf{x}$. $|x|$ indicates the absolute value of x, and $(x)_+ \triangleq \max(0, x)$. sign$(\cdot)$ is the component-wise sign function. $\mathbf{I}$ and $\mathbf{0}$ refer to the identity matrix and zero vector/matrix (their sizes are determined from the context). Finally, diag$(\mathbf{x})$ is the diagonal matrix whose i-th diagonal element is the i-th entry of $\mathbf{x}$.

II. BACKGROUND AND PROBLEM FORMULATION

A. SPARSITY-REGULARIZED STAP

Consider a radar platform equipped with a uniform linear array (ULA) consisting of N receiving elements, and the speed of radar is v_p. K pulses are transmitted at a constant pulse repetition frequency f_r in a coherent processing interval (CPI). The received signal from the range bin of interest is represented by [23]

$$\mathbf{x} = \sum_{n=1}^{N_c} \sigma_{c,n} \mathbf{v}\left(f_{d,n}, f_{s,n}\right) + \mathbf{n}, \tag{1}$$

where N_c and $\sigma_{c,n}$ denote the number and the complex reflection coefficient of clutter patches, respectively. $f_{d,n}$ and $f_{s,n}$ are the Doppler frequency and spatial frequency

for the nth clutter patch, respectively. $\mathbf{v}\left(f_{d,n}, f_{s,n}\right)$ is the space-time steering vector, which is defined as a Kronecker product of the temporal and spatial steering vectors [24], i.e., $\mathbf{v}\left(f_{d,n}, f_{s,n}\right) = \mathbf{v}_d\left(f_{d,n}\right) \otimes \mathbf{v}_s\left(f_{s,n}\right)$, where

$$\begin{aligned} \mathbf{v}_d\left(f_{d,n}\right) &= \left[\begin{array}{cccc} 1 & e^{j2\pi f_{d,n}} & \cdots & e^{j2\pi(K-1)f_{d,n}} \end{array}\right]^{\mathrm{T}} \\ \mathbf{v}_s\left(f_{s,n}\right) &= \left[\begin{array}{cccc} 1 & e^{j2\pi f_{s,n}} & \cdots & e^{j2\pi(N-1)f_{s,n}} \end{array}\right]^{\mathrm{T}}. \end{aligned} \tag{2}$$

$\mathbf{n}$ is the thermal noise vector.

The generalized side-lobe canceler (GSC) form of STAP algorithm is presented in Fig. 2. $\mathbf{v}_t$ is the steering vector of the moving target, and $\mathbf{B}$ is the feature matrix, whose columns can be regarded as bases of clutter subspace. The feature matrix transforms the original signal data into a feature space, and it is also called signal blocking matrix and satisfies $\mathbf{B}^{\mathrm{H}}\mathbf{v}_t = \mathbf{0}$ and $\mathbf{B}\mathbf{B}^{\mathrm{H}} = \mathbf{I}$.

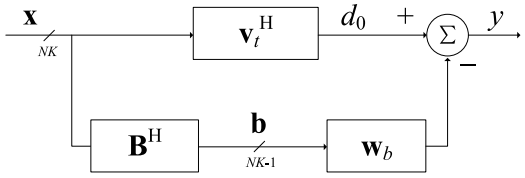


FIGURE 2. GSC form of the STAP.

After the transformation by $\mathbf{b} = \mathbf{B}^{\mathrm{H}}\mathbf{x}$, $NK - 1$ data are available to suppress the clutter. The output of the GSC-STAP is

$$y = d_0 - \mathbf{w}_b^{\mathrm{H}}\mathbf{b}. \tag{3}$$

The aim is to minimize $|y|^2$, which can be rewritten as

$$\min_{\mathbf{w}_b} \mathbf{w}_b^{\mathrm{H}}\mathbf{R}_b\mathbf{w}_b - \mathbf{w}_b^{\mathrm{H}}\mathbf{r}_b - \mathbf{r}_b^{\mathrm{H}}\mathbf{w}_b, \tag{4}$$

where $\mathbf{R}_b$ is the clutter-plus-noise covariance matrix, $\mathbf{r}_b$ is the cross-correlation vector between d_0 and $\mathbf{b}$, and $d_0 = \mathbf{v}_t^{\mathrm{H}}\mathbf{x}$. Consequently, the weight vector is

$$\mathbf{w}_b = \mathbf{R}_b^{-1}\mathbf{r}_b, \tag{5}$$

In practice, $\mathbf{R}_b$ and $\mathbf{r}_b$ are unknown. Normally, they can be estimated as $\mathbf{R}_b = \sum_{l=1}^{L} \mathbf{b}(l)\,\mathbf{b}^{\mathrm{H}}(l)/L$ and $\mathbf{r}_b = \sum_{l=1}^{L} \mathbf{b}(l)\,d_0^*(l)/L$, respectively, where L denotes the number of snapshots. However, a large number of IID training snapshots are required to achieve a satisfactory performance in this case.

In fact, the columns in $\mathbf{B}$ can be regarded as bases of clutter-plus-jamming subspace. It is known that the rank of clutter-plus-jamming subspace is far less than system DoFs, therefore there is no need to use all the columns in $\mathbf{B}$ to construct the subspace. In other words, the weight vector should be sparse, and the minimization problem can be expressed as a least absolute shrinkage and selection operator (LASSO) problem [13]:

$$\min_{\mathbf{w}_b} \; -\mathbf{r}_b^{\mathrm{H}}\mathbf{w}_b - \mathbf{w}_b^{\mathrm{H}}\mathbf{r}_b + \mathbf{w}_b^{\mathrm{H}}\mathbf{R}_b\mathbf{w}_b + \lambda\|\mathbf{w}_b\|_1, \tag{6}$$

(6) is convex, and can be solved by the interior point method (IPM). The sparsity-regularized constraint moves most elements in $\tilde{\mathbf{w}}_b$ toward zero, and only few nonzero elements are selected as the best auxiliary data for clutter suppression. Consequently, the number of DoFs is reduced, and the snapshot limitation for estimating the clutter covariance matrix can be relaxed.

Nevertheless, the weight vectors corresponding to different range bins are assumed to be independent with each other. Indeed, the performance by simply restraining the sparsity of weight vector can be improved further if we learn the STAP tasks jointly compared with learning them independently.

B. SPATIAL SMOOTHNESS

There could exist spatial relationship among related STAP tasks. For example, if the radar system works in the squint mode, then the normalized Doppler frequency for the clutter patch with the angle of arrival (AOA) ψ is

$$f_d = \frac{2v_p}{\lambda f_r}\cos(\psi) = \frac{2v_p}{\lambda f_r}\cos(\theta+\alpha)\cos(\varphi), \tag{7}$$

where λ represents the wavelength. θ, α, and φ represent the azimuth angle, the squint angle, and the elevation angle of the clutter patch, respectively. Because the slant ranges of range bins are different, the elevation angles are different as well. Hence, the clutter environment is nonhomogeneous, and clutter powers of different range bins are distributed in different areas, as shown in Fig. 3.

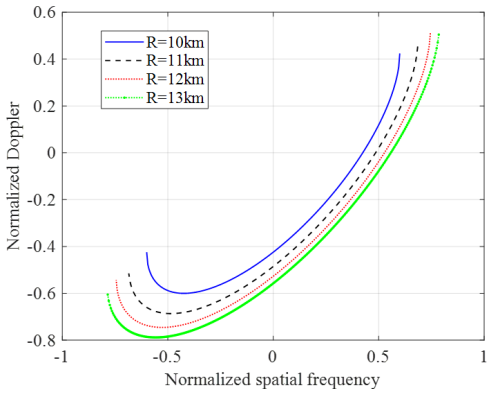


FIGURE 3. The clutter power distributions of different range bins. *R* denotes the slant range of range bin.

Although the clutter environment is nonhomogeneous, the weight vectors corresponding to neighboring STAP tasks can be similar and do not differ too much. To illustrate this, a simulation is given in Fig. 4. The weight vector corresponding to each range bin is trained by its ideal CCM. It can be seen that the deviation increases smoothly as spacing distance increases, i.e., there exists a spatial smoothness among related STAP tasks.

C. GROUP SPARSITY

In existing sparsity-regularized STAP algorithms, the sparsity commonality of weight vectors among related STAP

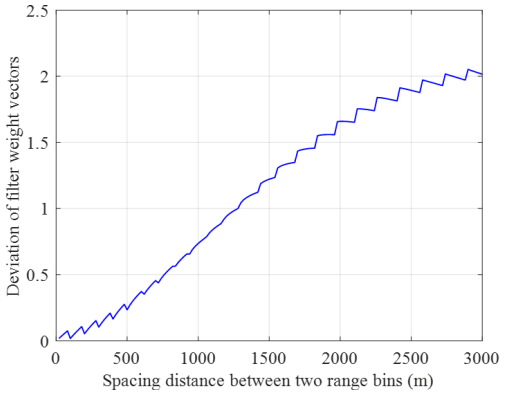


FIGURE 4. The deviation of weight vectors versus the spacing distance between two range bins. The deviation of weight vectors is defined as $\|\mathbf{w}_1 - \mathbf{w}_2\|_2^2$.

tasks has not been considered. Nevertheless, they are actually associated with each other. An example is illustrated in Fig. 5, which shows the selection results of auxiliary beams in different sparsity-regularised beam-space (BS) STAP tasks. In the BS-STAP task, $\mathbf{v}_t$ is regarded as the beam of interest in the angular-Doppler beam domain, and all other angular-Doppler beams are used to form the feature matrix $\mathbf{B}$ [9]. In RD BS-STAP algorithms, only a small portion of beams (i.e., auxiliary beams) are required to suppress the clutter and jamming. Apparently, in Fig. 5(a) and Fig. 5(b), the distributions of clutter/jamming in the angular-Doppler domain are not the same. Whereas, even though the clutter/jamming environments are nonhomogeneous in two STAP tasks, it can be seen that the auxiliary beams selected by sparsity-regularised STAP algorithm are almost the same, i.e., different STAP tasks share the same sparsity commonality.

III. MULTI-TASK LEARNING STAP

Motivated by the significant spatial smoothness and sparsity commonality of weight vectors, we propose a multi-task STAP algorithm by ensuring that the weight vectors corresponding to neighboring range bins are consistent and constraining all weight vectors to share a common feature selection. Explicitly, the problem is formulated as

$$\min_{\mathbf{W}_b} L\left(\mathbf{W}_b\right) + \lambda \sum_{t=1}^{T-1} \left\|\mathbf{w}_{b,t} - \mathbf{w}_{b,t+1}\right\|_2^2 + \gamma \left\|\mathbf{W}_b\right\|_{2,1}^2, \quad (8)$$

where $\mathbf{W}_b = \left[\mathbf{w}_{b,1}, \mathbf{w}_{b,2}, \cdots, \mathbf{w}_{b,T}\right]$ denotes the weight matrix, $L\left(\mathbf{W}_b\right) = \sum_{t=1}^{T}(\mathbf{w}_{b,t}^{\mathrm{H}}\mathbf{R}_{b,t}\mathbf{w}_{b,t} - \mathbf{r}_{b,t}^{\mathrm{H}}\mathbf{w}_{b,t} - \mathbf{w}_{b,t}^{\mathrm{H}}\mathbf{r}_{b,t})$ denotes the total mean square loss, and T denotes the number of STAP tasks. $\mathbf{w}_{b,t}$, $\mathbf{R}_{b,t}$ and $\mathbf{r}_{b,t}$ denote the weight vector, the covariance matrix and the cross-correlation vector of the t-th task, respectively. The second term, i.e., spatial smoothness term, penalizes large deviations of weight vectors corresponding to neighboring range bins of interest. This term can be expressed as

$$\sum_{t=1}^{T-1} \left\|\mathbf{w}_{b,t} - \mathbf{w}_{b,t+1}\right\|_2^2 = \left\|\mathbf{W}_b\mathbf{H}\right\|_F^2, \quad (9)$$

where $\mathbf{H}$ is defined by

$$H_{i,j} = \begin{cases} 1, & i = j \\ -1, & i = j+1 \\ 0, & otherwise. \end{cases} \quad (10)$$

The third term encourages a row-wise sparsity in $\mathbf{W}_b$, which is equivalent to a common feature selection. λ and γ control the spatial smoothness and group sparsity regularizations, respectively.

The objective in (8) can be considered as a combination of a smooth term and a non-smooth term. The ADMM [25] can be applied to solve the optimization. The variable $\mathbf{W}_b$ is split into a pair of variables $\mathbf{W}_b$ and $\mathbf{Z}$, and the two variables are equal:

$$\begin{aligned} &\min L\left(\mathbf{W}_b\right) + \lambda \left\|\mathbf{W}_b\mathbf{H}\right\|_F^2 + \gamma \left\|\mathbf{Z}\right\|_{2,1}^2 \\ &\text{s.t. } \mathbf{W}_b = \mathbf{Z}. \end{aligned} \quad (11)$$

As in the ADMM method, the augmented Lagrangian function is written as

$$\begin{aligned} \mathcal{L}_\rho = L\left(\mathbf{W}_b\right) &+ \lambda \left\|\mathbf{W}_b\mathbf{H}\right\|_F^2 + \gamma \left\|\mathbf{Z}\right\|_{2,1}^2 \\ &+ \langle \Sigma, \mathbf{W}_b - \mathbf{Z}\rangle + \frac{\rho}{2}\left\|\mathbf{W}_b - \mathbf{Z}\right\|_F^2, \end{aligned} \quad (12)$$

where $\rho > 0$ is the augmented Lagrangian parameter determining the penalty for the equality violation, and Σ is the Lagrange multiplier matrix. Define the residual and the scaled dual variable as $\mathbf{R} = \mathbf{W}_b - \mathbf{Z}$ and $\mathbf{D} = \frac{1}{\rho}\Sigma$, respectively. Then, we have

$$\begin{aligned} \mathcal{L}_\rho = L\left(\mathbf{W}_b\right) &+ \lambda \left\|\mathbf{W}_b\mathbf{H}\right\|_F^2 + \gamma \left\|\mathbf{Z}\right\|_{2,1}^2 \\ &+ \frac{\rho}{2}\left\|\mathbf{R} + \mathbf{D}\right\|_F^2 - \frac{\rho}{2}\left\|\mathbf{D}\right\|_F^2. \end{aligned} \quad (13)$$

Subsequently, the solution can be rewritten in the following form:

$$\begin{cases} \mathbf{W}_b^{(k+1)} = \arg\min\limits_{\mathbf{W}_b} L\left(\mathbf{W}_b\right) + \lambda \left\|\mathbf{W}_b\mathbf{H}\right\|_F^2 \\ \quad + \frac{\rho}{2}\left\|\mathbf{W}_b - \mathbf{Z}^{(k)} + \mathbf{D}^{(k)}\right\|_F^2 \\ \mathbf{Z}^{(k+1)} = \arg\min\limits_{\mathbf{Z}} \gamma \left\|\mathbf{Z}\right\|_{2,1}^2 + \frac{\rho}{2}\left\|\mathbf{W}_b^{(k+1)} - \mathbf{Z} + \mathbf{D}^{(k)}\right\|_2^2 \\ \mathbf{D}^{(k+1)} = \mathbf{D}^{(k)} + \mathbf{R}^{(k+1)}, \end{cases} \quad (14)$$

where $\mathbf{R}^{(k)} = \mathbf{W}_b{}^{(k)} - \mathbf{Z}^{(k)}$ is the residual at the kth iteration, and $\mathbf{D}^{(k)} = \mathbf{D}^{(0)} + \sum_{j=1}^{k}\mathbf{R}^{(j)}$ is the summation of the residuals.

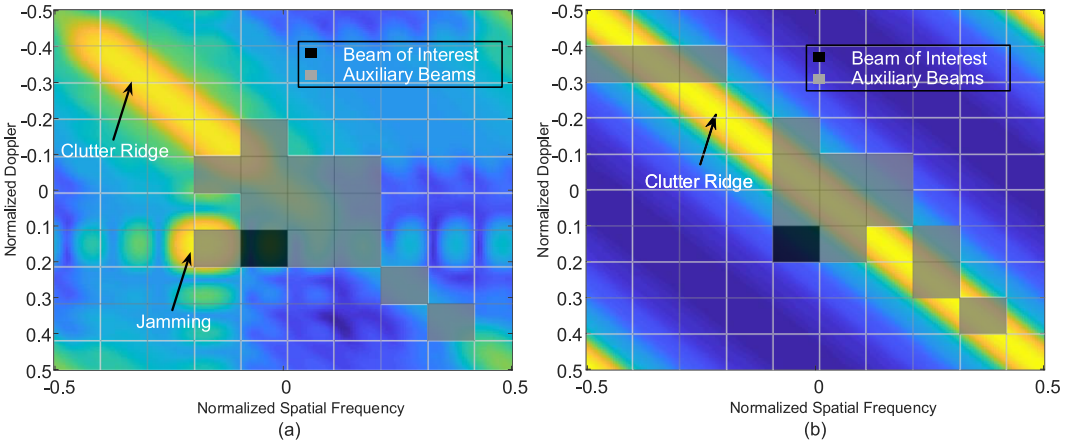


FIGURE 5. Selections of auxiliary beams under different circumstances. The black bins represent the beam of interest, and the grey ones represent the selected auxiliary beams for clutter/jamming suppression.

A. UPDATING $\mathbf{W}_b$ WITH FIXED VARIABLES $\mathbf{Z}$ AND $\mathbf{D}$

The first optimization problem in (14) has an analytical solution, as shown below. First, the first optimization problem in (14) can be rewritten as a quadratic function:

$$\begin{aligned}\min_{\mathbf{w}_b} \mathbf{w}_b^{\mathrm{H}} \left(\mathbf{R}_b + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \frac{\rho}{2}\mathbf{I}_{T\times(NK-1)}\right)\mathbf{w}_b \\ - \left[\mathbf{r}_b^{\mathrm{H}} + \frac{\rho}{2}\left(\mathbf{z}^{(k)} - \mathbf{d}^{(k)}\right)^{\mathrm{H}}\right]\mathbf{w}_b \\ - \mathbf{w}_b^{\mathrm{H}}\left[\mathbf{r}_b + \frac{\rho}{2}\left(\mathbf{z}^{(k)} - \mathbf{d}^{(k)}\right)\right],\end{aligned} \tag{15}$$

where

$$\mathbf{R}_b = \begin{bmatrix} \mathbf{R}_{b,1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{b,T} \end{bmatrix}, \quad \mathbf{r}_b = \begin{bmatrix} \mathbf{r}_{b,1} \\ \vdots \\ \mathbf{r}_{b,T} \end{bmatrix}, \tag{16}$$

and $\tilde{\mathbf{H}} = \mathbf{H} \otimes \mathbf{I}_{NK-1}$. $\mathbf{w}_b$, $\mathbf{z}$ and $\mathbf{d}$ are the vectorization operations of $\mathbf{W}_b$, $\mathbf{Z}$ and $\mathbf{D}$, respectively. Hence, the objective is to minimize a quadratic function, and the solution can be directly obtained as

$$\mathbf{w}_b = \left(\mathbf{R}_b + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \frac{\rho}{2}\mathbf{I}\right)^{-1}\left[\mathbf{r}_b + \frac{\rho}{2}\left(\mathbf{z}^{(k)} - \mathbf{d}^{(k)}\right)\right]. \tag{17}$$

However, the inversion of $\left(\mathbf{R}_b + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \frac{\rho}{2}\mathbf{I}\right)$ has a high computational complexity of $\mathcal{O}(T^3\,(NK-1)^3)$, especially in the case that the number of tasks is large. Whereas, the computational complexity can be reduced to $\mathcal{O}((2T-1)\,(NK-1)^3)$, and a detailed method is presented in APPENDIX A. Note that T tasks are solved simultaneously, hence the computational complexity is $\mathcal{O}\left(\frac{2T-1}{T}(NK-1)^3\right)$ for each STAP task.

B. UPDATING $\mathbf{Z}$ WITH FIXED VARIABLES $\mathbf{W}_b$ AND $\mathbf{D}$

The second optimization problem in (14) can be decomposed into $NK-1$ independent optimization problems, which can be rewritten as

$$\min_{\mathbf{Z}} \sum_{d=1}^{NK-1} (\gamma\|\mathbf{z}_d\|_2 + \frac{\rho}{2}\left\|\mathbf{w}_{b,d}^{(k+1)} - \mathbf{z}_d + \mathbf{d}_d^{(k)}\right\|_2^2), \tag{18}$$

where $\mathbf{z}_d$, $\mathbf{w}_{b,d}$ and $\mathbf{d}_d$ are the d-th row of $\mathbf{Z}$, $\mathbf{W}_b$ and $\mathbf{D}$, respectively. The closed-form solution is given by

$$\mathbf{z}_d^{(k+1)} = \left(1 - \frac{\gamma}{\rho\left\|\mathbf{w}_{b,d}^{(k+1)} + \mathbf{d}_d^{(k)}\right\|_2}\right)_+ (\mathbf{w}_{b,d}^{(k+1)} + \mathbf{d}_d^{(k)}). \tag{19}$$

We summarize the proposed algorithm in Algorithm 1. A proof of the convergence is presented in APPENDIX B.

Algorithm 1 Multi-Task Learning STAP Algorithm

Input: training snapshot sets $\{\mathbf{X}_t\}, t = 1, \cdots, T$
Parameters: regularization parameters λ, γ, ρ, and stopping threshold *tol*
Output: weight matrix $\mathbf{W}_b$
Initialization: set $\mathbf{Z} = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$
Compute $\left(\mathbf{R}_b + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \frac{\rho}{2}\mathbf{I}\right)^{-1}$ according to (34)
while $\left\|\mathbf{W}_b^{(k+1)} - \mathbf{W}_b^{(k)}\right\|_F^2 \geq tol$ **do**
1. Update $\mathbf{W}_b$ according to (15)
2. Update $\mathbf{Z}$ according to (19)
3. Update $\mathbf{D}$ via $\mathbf{D}^{(k+1)} = \mathbf{D}^{(k)} + \mathbf{R}^{(k+1)}$

end while

IV. MULTI-TASK FEATURE LEARNING STAP

In our previous discussions, the feature matrix $\mathbf{B}$ is assumed as prescribed or known. For example, $\mathbf{B}$ can be obtained by singular value decomposition [12], or formed by angular-Doppler beams [9]. However, these prescribed feature matrices are not necessarily the optimal ones for the multi-task learning STAP, and it might be better to learn a feature matrix so that the associations among STAP tasks can be enhanced.

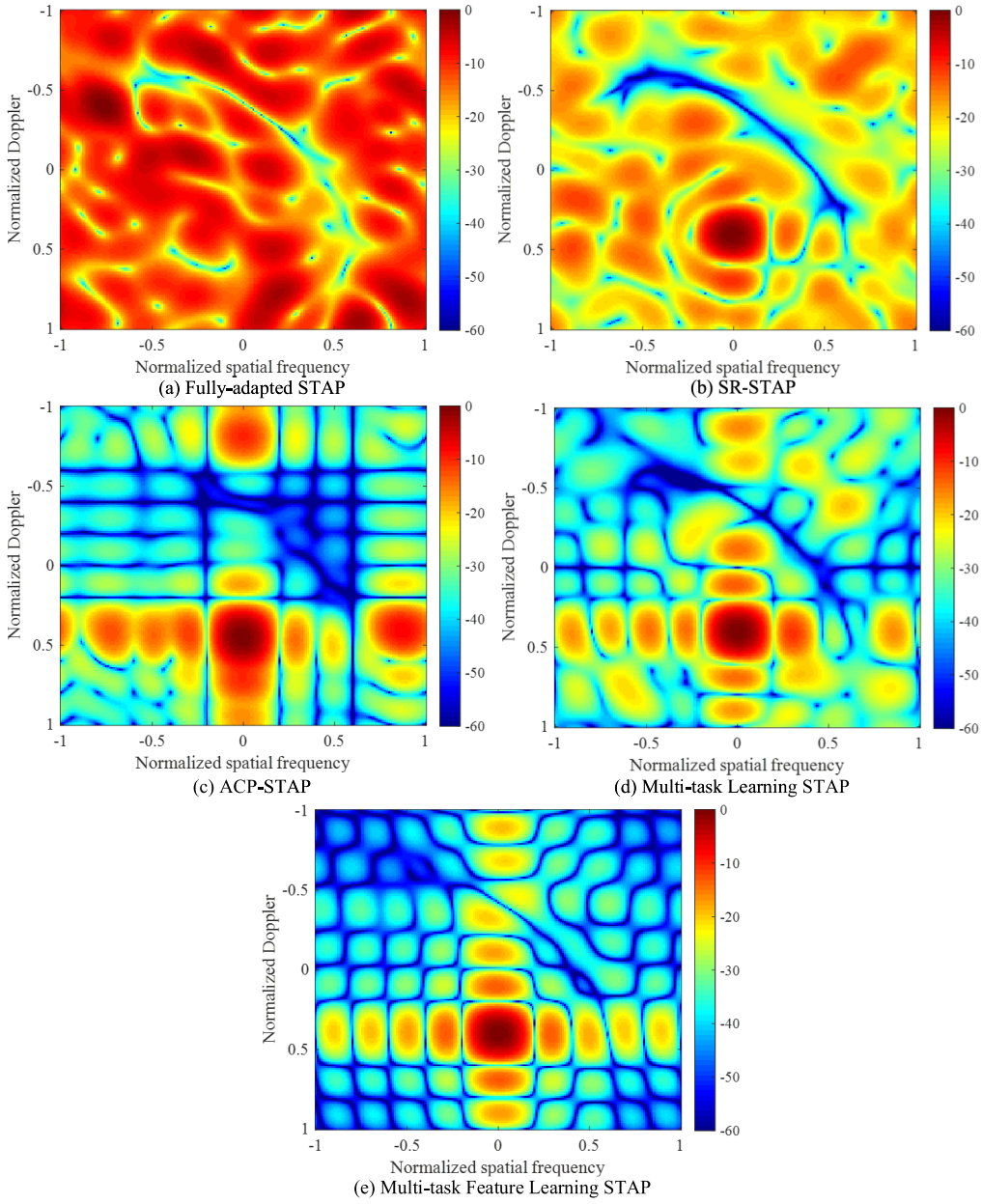


FIGURE 6. **Normalized beampattern versus normalized Doppler and spatial frequency.**

Thus, a multi-task feature learning STAP algorithm is proposed to jointly optimize $\mathbf{W}_b$ and $\mathbf{B}$:

$$\begin{aligned}\min_{\mathbf{W_b},\mathbf{B}} & L\left(\mathbf{W}_b,\mathbf{B}\right)+\lambda\sum_{t=1}^{T-1}\left\|\mathbf{B}\mathbf{w}_{b,t}-\mathbf{B}\mathbf{w}_{b,t+1}\right\|_2^2 \\ & +\gamma\left\|\mathbf{W}_b\right\|_{2,1}^2 \\ \text{s.t. } & \mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I},\end{aligned} \tag{20}$$

where $L\left(\mathbf{W}_b,\mathbf{B}\right)$ equals to

$$\sum_{t=1}^{T}\left(\mathbf{w}_{b,t}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{R}_{x,t}\mathbf{B}\mathbf{w}_{b,t}-\mathbf{r}_t^{\mathrm{H}}\mathbf{B}\mathbf{w}_{b,t}-\mathbf{w}_{b,t}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{r}_t\right). \tag{21}$$

$\mathbf{R}_{x,t}=E\left(\mathbf{x}_t\mathbf{x}_t{}^{\mathrm{H}}\right)$ and $\mathbf{r}_t=E\left(\mathbf{x}_t d_{t,0}^*\right)$ are the input covariance matrix and cross-correlation vector of the t-th STAP task, respectively.

The optimization problem (20) is challenging. Different from the convex problem in (8), it is non-convex. However, inspired by the method in [26], an equivalent convex optimization problem can be obtained:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{V}} & L\left(\mathbf{W}\right)+\lambda\left\|\mathbf{W}\mathbf{H}\right\|_F^2+\gamma\sum_{t=1}^{T}\mathbf{w}_t^{\mathrm{H}}\mathbf{V}^{-1}\mathbf{w}_t \\ \text{s.t. } & \mathbf{V}\succeq 0,\quad \mathrm{trace}(\mathbf{V})\le 1,\end{aligned} \tag{22}$$

Algorithm 2 Multi-Task Feature Learning STAP Algorithm

Input: training snapshot sets $\{\mathbf{X}_t\}, t = 1, \cdots, T$
Parameters: regularization parameters λ, γ, and stopping threshold *tol*
Output: weight matrix $\mathbf{W}$
Initialization: set $\mathbf{V} = \frac{\mathbf{I}_{NK}}{NK}$
while $\left\|\mathbf{W}^{(k+1)} - \mathbf{W}^{(k)}\right\|_F^2 \geq tol$ **do**
1. Compute $\left(\mathbf{R}_x + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \gamma\tilde{\mathbf{V}}^{-1}\right)^{-1}$ via the acceleration method in APPENDIX A
2. Update $\mathbf{W}$ according to (25)
3. Update $\mathbf{V}$ according to (29)
end while

where $L(\mathbf{W}) = \sum_{t=1}^{T}(\mathbf{w}_t^{\mathrm{H}}\mathbf{R}_{x,t}\mathbf{w}_t - \mathbf{r}_t^{\mathrm{H}}\mathbf{w}_t - \mathbf{w}_t^{\mathrm{H}}\mathbf{r}_t)$. In particular, if $(\mathbf{W}_b, \mathbf{B})$ is an optimal solution of (20), then

$$(\mathbf{W}, \mathbf{V}) = \left(\mathbf{B}\mathbf{W}_b, \mathbf{B}\mathrm{diag}\left(\frac{\|\mathbf{w}_{b,d}\|_2}{\|\mathbf{W}_b\|_{2,1}}\right)_{d=1}^{NK-1}\mathbf{B}^{\mathrm{H}}\right) \quad (23)$$

is an optimal solution of (22). The proof of equivalence can be seen in [26].

The problem can be solved by alternating minimization algorithm. In the first step, we keep $\mathbf{V}$ fixed and minimize over $\mathbf{W}$, i.e., we solve the unconstrained problem

$$\min_{\mathbf{W}} L(\mathbf{W}) + \lambda\|\mathbf{W}\mathbf{H}\|_F^2 + \gamma\sum_{t=1}^{T}\mathbf{w}_t^{\mathrm{H}}\mathbf{V}^{-1}\mathbf{w}_t, \quad (24)$$

which can be rewritten as a quadratic problem:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\left(\mathbf{R}_x + \lambda\tilde{\mathbf{H}}^{\mathrm{H}} + \gamma\tilde{\mathbf{V}}^{-1}\right)\mathbf{w} - \mathbf{r}^{\mathrm{H}}\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{r}, \quad (25)$$

where

$$\mathbf{R}_x = \begin{bmatrix} \mathbf{R}_{x,1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{x,T} \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_T \end{bmatrix}, \quad (26)$$

and $\tilde{\mathbf{V}}^{-1} = \mathbf{I}_T \otimes \mathbf{V}^{-1}$. Hence, the solution can be directly obtained as

$$\mathbf{w} = \left(\mathbf{R}_x + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \gamma\tilde{\mathbf{V}}^{-1}\right)^{-1}\mathbf{r}. \quad (27)$$

Similarly, the inversion operation can be accelerated by the method in APPENDIX A.

In the second step, we keep $\mathbf{W}$ fixed and minimize over $\mathbf{V}$, i.e., we solve the problem

$$\min_{\mathbf{W}} \sum_{t=1}^{T}\mathbf{w}_t^{\mathrm{H}}\mathbf{V}^{-1}\mathbf{w}_t \quad \text{s.t. } \mathbf{V} \succeq 0, \quad \mathrm{trace}(\mathbf{V}) \leq 1. \quad (28)$$

The analytic solution is given by

$$\mathbf{V} = \frac{\left(\mathbf{W}\mathbf{W}^{\mathrm{H}}\right)^{\frac{1}{2}}}{\mathrm{trace}\left(\mathbf{W}\mathbf{W}^{\mathrm{H}}\right)^{\frac{1}{2}}}. \quad (29)$$

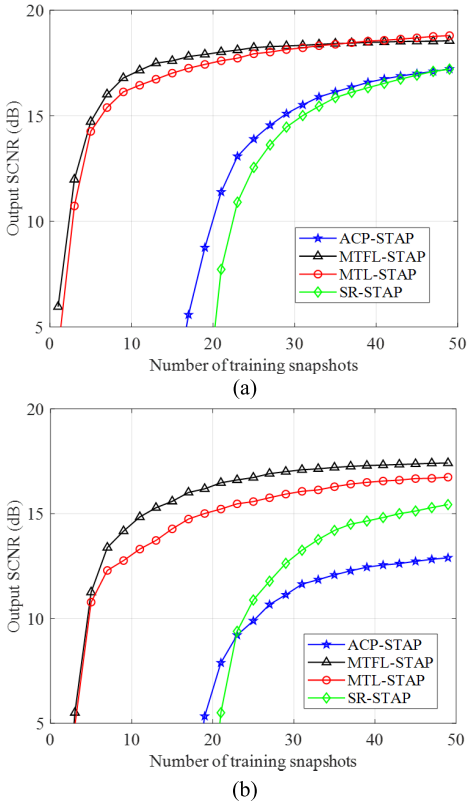


FIGURE 7. Output SCNR versus the number of training snapshots. (a) The normalized doppler frequency of the moving target is 0.4 (b) The normalized doppler frequency of the moving target is −0.2.

We summarize the proposed multi-task feature learning STAP algorithm in Algorithm 2. Algorithm 2 can be interpreted as two steps. In the first step, we learn task-specific functions using a common representation across the tasks. In the second step, we learn the common representation.

V. SIMULATION RESULTS

In this section, the performance of the proposed algorithm is verified by numerical simulations. Consider an airborne radar system equipped with $N = 10$ receiving elements, and the elements are spaced half a wavelength apart, i.e., $d = \lambda/2$. The radar works in large squint mode where the squint angle is $\alpha = 45°$, hence the clutter environment is nonhomogeneous. The radar transmits $K = 10$ pulses during a CPI, and the flight height H is 8000m. Additive noise is modeled as spatially and temporally independent complex Gaussian noise. Let $f_r = 4v_p\lambda$, hence, $2v_pT_r/d = 1$. The signal of the moving target impinges the array from a AOA of $\psi_t = 0°$, and the signal-to-noise ratio is 0dB.

To demonstrate the superior performance of the proposed algorithm, some popular algorithms including the conventional fully-adapted STAP, SR-STAP [13] and ACP-STAP [27] are compared in the following simulations. In the multi-task learning STAP, $T = 5$ STAP tasks are jointly optimized, and the t-th STAP task aims to obtain a weight vector for the clutter suppression when we detect the moving

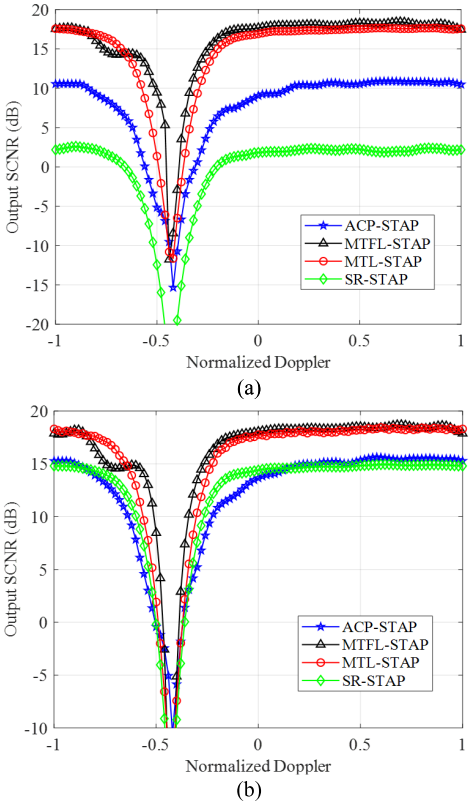


FIGURE 8. Output SCNR versus the target Doppler frequency. (a) The number of training snapshots is 20 (b) The number of training snapshots is 30.

target in the t-th range bin of interest. The slant ranges of T range bins of interest are 10km, 10.5km, 11km, 11.5km and 12km, respectively. The sparsity regularization parameters λ and γ are set to 1000 and 1, respectively. All the results are obtained from the average of 100 independent Monte-Carlo simulations.

A. BEAMPATTERNS IN ANGLE-DOPPLER

The normalized beampatterns of different STAP algorithms are shown in Fig. 6. The normalized Doppler frequency of the moving target $f_{d,t}$ is set to 0.4, and 10 training snapshots for each STAP task is used to train the weight vector in a small number of samples. First, as shown in Fig. 6(a), the fully-adapted STAP algorithm cannot work normally due to the lack of sufficient training snapshots to estimate the CCM precisely. The sidelobe level is high, and there is no response peak value and deep null in the areas of target and clutter, respectively. Second, as shown in Fig. 6(b) and Fig. 6(c), although a deep null is formed in the area of clutter, but the sidelobe level is still high, which increases the false alarm probability and deteriorates the output SCNR performance. Moreover, as shown in Fig. 6(d) and Fig. 6(e), the sidelobe levels of the proposed MTL-STAP and MTFL-STAP algorithms are low, and the mainlobes are capable of correctly pointing towards the target. Also, the deep nulls are placed in the area of clutter. This means that the proposed algorithms can extract useful information in a small number of samples to estimate the clutter-plus-noise power spectrum accurately.

B. OUTPUT SCNR PERFORMANCE

The output SCNR performances versus the number of training snapshots and the target Doppler frequency are compared in Fig. 7 and Fig. 8. As shown in these figures, we can see that: (i) the output SCNR performances of the proposed MTL-STAP and MTFL-STAP algorithms are superior to that of the SR-STAP and ACP-STAP algorithms. Due to the fact that the correlated information among different STAP tasks can be used to improve the learning of each task, MTL-STAP and MTFL-STAP algorithms can reliably estimate the filter parameters within a few snapshots; (ii) the output SCNR performance of the MTFL-STAP algorithm can outperform that of the MTL-STAP algorithm in some cases, which supports the conclusion that the prescribed feature matrix is not necessarily the optimal one, and it might be better to learn a feature matrix simultaneously to enhance the associations among STAP tasks.

VI. CONCLUSION

Due to the strong associations among STAP tasks, in this paper, we proposed a novel multi-task learning STAP method via spatial smoothness regularization to accelerate the convergence speed of STAP, and the problem was solved by ADMM effectively. Furthermore, in the case where the feature matrix was unknown, a multi-task feature learning STAP method was proposed to jointly optimize the feature matrix and the weight matrix. The numerical results demonstrated that the proposed methods with the shared information among multiple related STAP tasks can effectively decrease the required number of training snapshots and provide a better performance.

APPENDIX A
ACCELERATION METHOD FOR THE INVERSION OPERATION

Denote $\tilde{\mathbf{R}}_0 = \mathbf{R}_b + \frac{\rho}{2}\mathbf{I}$. Note that

$$\tilde{\mathbf{R}}_0 = \begin{bmatrix} \mathbf{R}_{b,1} + \frac{\rho}{2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{b,T} + \frac{\rho}{2}\mathbf{I} \end{bmatrix}. \tag{30}$$

Hence,

$$\tilde{\mathbf{R}}_0^{-1} = \begin{bmatrix} \left(\mathbf{R}_{b,1} + \frac{\rho}{2}\mathbf{I}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{R}_{b,T} + \frac{\rho}{2}\mathbf{I}\right)^{-1} \end{bmatrix}, \tag{31}$$

which has a computational complexity of $\mathcal{O}(T\,(NK-1)^3)$.

Then, note that $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}$ can be rewritten as

$$\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} = \sum_{t=1}^{T-1} \mathbf{U}_t \mathbf{U}_t^{\mathrm{H}}, \tag{32}$$

where

$$\mathbf{U}_t = \begin{bmatrix} \mathbf{0}_{t-1} \\ 1 \\ -1 \\ \mathbf{0}_{T-t-1} \end{bmatrix} \otimes \mathbf{I}_{NK-1} \in \mathbb{C}^{T(NK-1)\times(NK-1)}. \quad (33)$$

Denote $\tilde{\mathbf{R}}_t = \tilde{\mathbf{R}}_0 + \lambda \sum_{\tilde{t}=1}^{t} \mathbf{U}_{\tilde{t}}\mathbf{U}_{\tilde{t}}^{\mathrm{H}}$, then it is clear that $\tilde{\mathbf{R}}_t = \tilde{\mathbf{R}}_{t-1} + \lambda \mathbf{U}_t\mathbf{U}_t^{\mathrm{H}}$. According to the matrix inversion lemma [28], we obtain

$$\tilde{\mathbf{R}}_t^{-1} = \tilde{\mathbf{R}}_{t-1}^{-1} - \tilde{\mathbf{R}}_{t-1}^{-1}\mathbf{U}_t\left(\mathbf{U}_t^{\mathrm{H}}\tilde{\mathbf{R}}_{t-1}^{-1}\mathbf{U}_t + \frac{1}{\lambda}\mathbf{I}_{NK-1}\right)^{-1}\mathbf{U}_t^{\mathrm{H}}\tilde{\mathbf{R}}_{t-1}^{-1}. \quad (34)$$

Through $T-1$ iterations, $\tilde{\mathbf{R}}_{T-1}^{-1}$ is obtained, which equals to $\left(\mathbf{R}_b + \lambda\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}} + \frac{\rho}{2}\mathbf{I}\right)^{-1}$. Because the matrix multiplication can be efficiently performed in parallel, the computational load is mainly determined by the inversion of $\mathbf{U}_t^{\mathrm{H}}\tilde{\mathbf{R}}_{t-1}^{-1}\mathbf{U}_t + \frac{1}{\lambda}\mathbf{I}_{NK-1}$. Hence, the computational complexity of iterations is $\mathcal{O}((T-1)(NK-1)^3)$.

Consequently, the total computational complexity is reduced from $\mathcal{O}(T^3(NK-1)^3)$ to $\mathcal{O}((2T-1)(NK-1)^3)$.

APPENDIX B
CONVERGENCE ANALYSIS OF ALGORITHM 1

We begin our proof by presenting the following theorem.

Theorem 1 (Eckstein-Bertsekas, [29]): Consider the problem

$$\min_{\mathbf{u}} f_1(\mathbf{u}) + f_2(\mathbf{v}) \quad s.t.\ \mathbf{v} = \mathbf{G}\mathbf{u} \quad (35)$$

in the case where the functions $f_1(\cdot)$ and $f_2(\cdot)$ are closed, proper, and convex, and $\mathbf{G}$ has a full column rank. Let $\{\eta_k \geq 0, k = 0, 1, \cdots\}$ and $\{\gamma_k \geq 0, k = 0, 1, \cdots\}$ be two sequences such that

$$\sum_{k=0}^{\infty} \eta_k < \infty \text{ and } \sum_{k=0}^{\infty} \gamma_k < \infty. \quad (36)$$

Assume that there are three sequences $\{\mathbf{u}_k, k = 0, 1, \cdots\}$, $\{\mathbf{v}_k, k = 0, 1, \cdots\}$, and $\{\mathbf{t}_k, k = 0, 1, \cdots\}$ that satisfy

$$\begin{aligned} \eta_k &\geq \left\| \mathbf{u}_{k+1} - \operatorname{argmin}_{\mathbf{u}} \left\{ f_1(\mathbf{u}) + (\rho/2)\|\mathbf{G}\mathbf{u} - \mathbf{v}_k - \mathbf{t}_k\|_2^2 \right\} \right\| \\ \gamma_k &\geq \left\| \mathbf{v}_{k+1} - \operatorname{argmin}_{\mathbf{v}} \left\{ f_2(\mathbf{v}) + (\rho/2)\|\mathbf{G}\mathbf{u}_{k+1} - \mathbf{v} - \mathbf{t}_k\|_2^2 \right\} \right\| \\ \mathbf{t}_{k+1} &= \mathbf{t}_k - (\mathbf{G}\mathbf{u}_{k+1} - \mathbf{v}_{k+1}) \end{aligned} \quad (37)$$

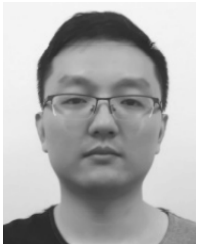
Then, if (35) has an optimal solution $\mathbf{u}^{\dagger}$, the sequence $\{\mathbf{u}_k\}$ converges to this solution, i.e., $\mathbf{u}_k \to \mathbf{u}^{\dagger}$.

First, since (11) is a particular instance when $\mathbf{G} = \mathbf{I}$, the full-rank condition in Theorem 1 can be satisfied. Second, it is clear that $f_1(\mathbf{W}_b) = L(\mathbf{W}_b) + \lambda\|\mathbf{W}_b\mathbf{H}\|_F^2$ and $f_2(\mathbf{Z}) = \gamma\|\mathbf{Z}\|_{2,1}^2$ in (11) are closed, proper, and convex. Moreover, the sequences $\mathbf{W}_b^{(k)}$, $\mathbf{Z}^{(k)}$, and $\mathbf{D}^{(k)}$ generated by (14) satisfy the conditions of (37) in a strict sense ($\eta_k = \gamma_k = 0$). Hence, the convergence is guaranteed.

REFERENCES

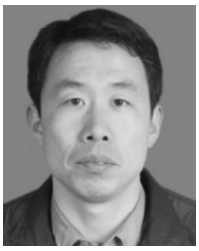
[1] W. L. Melvin, "A STAP overview," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 19–35, Jan. 2004.
[2] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA, USA: Artech House, 2014.
[3] B. Tang, J. Tuck, and P. Stoica, "Polyphase waveform design for MIMO radar space time adaptive processing," *IEEE Trans. Signal Process.*, vol. 68, pp. 2143–2154, 2020.
[4] B. Tang, J. Tang, and Y. Peng, "Clutter nulling performance of SMI in amplitude heterogeneous clutter environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1366–1373, Apr. 2013.
[5] R. Fa and R. C. De Lamare, "Reduced-rank STAP algorithms using joint iterative optimization of filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1668–1684, Jul. 2011.
[6] X. Wang, E. Aboutanios, and M. G. Amin, "Reduced-rank STAP for slow-moving target detection by antenna-pulse selection," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1156–1160, Aug. 2015.
[7] R. Li, J. Li, W. Zhang, and Z. He, "Reduced-dimension space-time adaptive processing based on angle-Doppler correlation coefficient," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–9, Sep. 2016.
[8] Z. Wei, Z. He, J. Li, and H. Liu, "Multiple-input–'multiple-output radar multistage multiple-beam beamspace reduced-dimension space-time adaptive processing," *IET Radar, Sonar Navigat.*, vol. 7, no. 3, pp. 295–303, Mar. 2013.
[9] W. Zhang, Z. He, J. Li, H. Liu, and Y. Sun, "A method for finding best channels in beam-space post-doppler reduced-dimension STAP," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 254–264, Jan. 2014.
[10] K. Sun, H. Meng, F. D. Lapierre, and X. Wang, "Registration-based compensation using sparse representation in conformal-array STAP," *Signal Process.*, vol. 91, no. 10, pp. 2268–2276, Oct. 2011.
[11] K. Sun, H. Meng, Y. Wang, and X. Wang, "Direct data domain STAP using sparse representation of clutter spectrum," *Signal Process.*, vol. 91, no. 9, pp. 2222–2236, Sep. 2011.
[12] Z. Yang, R. C. De Lamare, and X. Li, "L_1 regularized stap algorithm with a generalized sidelobe canceler architecture for airborne radar," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 674–686, Feb. 2012.
[13] Z. Gao, H. Tao, S. Zhu, and J. Zhao, "ℓ_1-regularised joint iterative optimisation space-time adaptive processing algorithm," *IET Radar, Sonar Navigat.*, vol. 10, no. 3, pp. 435–441, Mar. 2016.
[14] Z. Yang, X. Li, H. Wang, and L. Nie, "Sparsity-based space-time adaptive processing using complex-valued Homotopy technique for airborne radar," *IET Signal Process.*, vol. 8, no. 5, pp. 552–564, Jul. 2014.
[15] M. Shen, J. Wang, D. Wu, and D. Zhu, "An efficient moving target detection algorithm based on sparsity-aware spectrum estimation," *Sensors*, vol. 14, no. 9, pp. 17055–17067, Sep. 2014.
[16] B. Tang, Y. Zhang, J. Tang, and Y. Peng, "Close form maximum likelihood covariance matrix estimation under a knowledge-aided constraint," *IET Radar, Sonar Navigat.*, vol. 7, no. 8, pp. 904–913, Oct. 2013.
[17] B. Tang, J. Tang, and Y. Peng, "Performance of knowledge aided space time adaptive processing," *IET Radar, Sonar Navigat.*, vol. 5, no. 3, pp. 331–340, Apr. 2011.
[18] S. Bidon, O. Besson, and J.-Y. Tourneret, "Knowledge-aided STAP in heterogeneous clutter using a hierarchical Bayesian algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1863–1879, Jul. 2011.
[19] M. Liu, L. Zou, X. Yu, Y. Zhou, X. Wang, and B. Tang, "Knowledge aided covariance matrix estimation via Gaussian kernel function for airborne SR-STAP," *IEEE Access*, vol. 8, pp. 5970–5978, 2020.
[20] W. Melvin and J. Guerci, "Knowledge-aided signal processing: A new paradigm for radar and other advanced sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 983–996, Jul. 2006.
[21] Y. Zhang and Q. Yang, "An overview of multi-task learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 30–43, Jan. 2018.
[22] K.-H. Thung and C.-Y. Wee, "A brief review on multi-task learning," *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29705–29725, Nov. 2018.
[23] P. He, S. He, Z. Yang, and P. Huang, "An off-grid STAP algorithm based on local mesh splitting with bistatic radar system," *IEEE Signal Process. Lett.*, vol. 27, pp. 1355–1359, 2020.
[24] L. Pallotta, A. Farina, S. T. Smith, and G. Giunta, "Phase-only space-time adaptive processing," *IEEE Access*, vol. 9, pp. 147250–147263, 2021.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jul. 2011.
[26] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Mach. Learn.*, vol. 73, no. 3, pp. 243–272, Dec. 2008.
[27] R. Klemm, "Adaptive airborne MTI: An auxiliary channel approach," *IEE Proc. F, Commun., Radar Signal Process.*, vol. 134, no. 3, pp. 269–276, Jun. 1987.
[28] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, "Quantized kernel recursive least squares algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1484–1491, Sep. 2013.
[29] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 3, pp. 293–318, Jun. 1992.

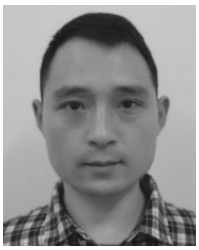


LILONG QIN was born in Anhui, China, in 1988. He received the M.S. degree in circuits and systems from the Electronic Engineering Institute, Hefei, China, in 2013, and the Ph.D. degree in communication and information systems engineering from the National University of Defense Technology, Changsha, China, in 2017. He is currently a Lecturer with the National University of Defense Technology. His research interests include synthetic aperture radar and adaptive beamforming.

BO TANG was born in Linchuan, Jiangxi, China, in 1985. He received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2006 and 2011, respectively. From July 2011 to June 2017, he was a Lecturer with the Electronic Engineering Institute. Since July 2017, he has been with the College of Electronic Engineering, National University of Defense Technology, China, where he is currently an Associate Professor. His research interests mainly include adaptive radar signal processing and radar waveform design.



HAI WANG was born in Shandong, China, in 1977. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Electronic Engineering Institute, Hefei, China, in 2000, 2003, and 2006, respectively. Since July 2017, he has been with the College of Electronic Engineering, National University of Defense Technology, China, where he is currently a Professor. His research interests mainly include data analysis and processing.



ZHONGRUI HUANG was born in Lankao, Henan, China, in 1988. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Electronic Engineering Institute, Hefei, China, in 2010, 2013, and 2016, respectively. Since July 2017, he has been with the College of Electronic Engineering, National University of Defense Technology, China, where he is currently a Lecturer. His research interests mainly include MIMO radar and array signal processing.