Article

Random Nyström Approach to Clutter Subspace Estimation for Polarimetric Space-Time Adaptive Processing

Kang Zhao, Zhiwen Liu, Shuli Shi, Yulin Huang and Yougen Xu *

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China;

k.zhao@bit.edu.cn (K.Z.); zwliu@bit.edu.cn (Z.L.); shishuli@bit.edu.cn (S.S.); yulinhuang@bit.edu.cn (Y.H.)

* Correspondence: yougenxu@bit.edu.cn; Tel.: +86-010-68912696

Received: 6 August 2020; Accepted: 8 September 2020; Published: 10 September 2020

Abstract: A random Nyström (R-Nyström) scheme for clutter subspace estimation is proposed in the context of polarimetric space-time adaptive processing (pSTAP). Unlike the standard Nyström scheme making use of only partial columns of the clutter plus noise covariance matrix (CNCM), R-Nyström exploits full CNCM information with a properly designed selection procedure under the newly developed random ridge cross leverage score (RRCLS) criterion. With R-Nyström, supported by the complete CNCM columns, upgraded clutter subspace estimation can be achieved at the expense of an insignificant increase in computational complexity, in contrast to the standard Nyström. The R-Nyström-based pSTAP, termed pR-Nyström, is shown to be superior over the current eigendecomposition-free subspace pSTAP in the signal to clutter plus noise loss and computational complexity. The efficacy of R-Nyström/pR-Nyström is validated by the simulation results.

Keywords: polarization; STAP; clutter suppression; clutter subspace; Nyström; ridge leverage score

1. Introduction

Polarimetric space-time adaptive processing (pSTAP) plays an important role in improving the detection performance of airborne array radar operating in clutter environments [1–4]. The purpose of pSTAP is to suppress clutter by maximizing the signal to clutter plus noise ratio (SCNR) [5] in the cell under test (CUT), with the direct use of the clutter plus noise covariance matrix (CNCM). In pSTAP, the estimation accuracy of CNCM determines the performance of the processor. In general, CNCM can be estimated by using the independent and identically distributed (i.i.d.) training samples obtained from the adjacent range bins of CUT. However, in a real nonhomogeneous environment, the clutter characteristic in different range cells would change quickly and dynamically, which caused difficulty in collecting enough training samples [6].

Subspace-based methods have also been studied in the literature by using the principal subspace of CNCM instead. For example, the eigen-canceller method based on the direct eigenvalue decomposition (EVD) of CNCM was developed in [7]. This method, however, is not attractive for real-time processing due to its high computational complexity. For this reason, some fast subspace methods without EVD of CNCM have been proposed further. For instance, the projection approximation subspace tracking (PAST) scheme in [8] and the fast approximated power iteration (FAPI) scheme in [9]. Both schemes are implemented iteratively for the principal subspace estimation with some training samples available. A variant of PAST, called L1-PAST, was developed in [10], and incorporates an additional sparse constraint on the filter. These mentioned EVD-free subspace methods are computationally more efficient than the EVD method but suffer a severe performance loss in the estimation of the principal subspace.

The Nyström scheme is a popular technique for the fast estimation of the matrix representation of the principal subspace [11–14] while it was originally designed for solving the integral equations [15].

The main idea of Nyström is to select a subset called landmark points for the calculation of the kernel eigenfunction approximately. The most favored sampling method for subset selection is random sampling, which leads to the fast versions of kernel machines [16] and spectral clustering [17].

Although the random sampling method can work in practice, it will fail for many natural kernel matrices where the relative “importance” of points is not uniform across the dataset. A ridge leverage score (RLS) method is proposed for measuring the importance of points [18–21] such that the kernel matrices can be selected more reasonably. When the number of selected columns is fixed, the corresponding number of rows of the selection matrix is also fixed. The performance of the existing Nyström-based subspace estimation method is proportional to the number of selected columns. Therefore, it is worth studying improving the estimation accuracy of the principal subspace without increasing the selection number [22].

Most current Nyström variants employ only partial column information of CNCM. There is still room to improve the application of Nyström for estimating the clutter subspace. In this paper, we propose a new method called random Nyström (R-Nyström) which makes full use of CNCM under the newly developed “random” ridge cross leverage score (RRCLS) rule. In the RRCLS rule, a new selection matrix is constructed for obtaining the low-dimensional matrix in Nyström. The R-Nyström scheme can be directly applied for implementing high computational efficient pSTAP.

The major contributions of the paper are the following:

- (1) Enabling fast principal subspace identification within the Nyström framework while making full use of the CNCM information.
- (2) The computational burden in pSTAP can be reduced with high clutter suppression performance compared with the existing methods.

2. Data Model

Consider a side-looking uniform linear polarimetric radar array system with N elements, each of which contains two independent polarimetric receiving channels. The inner-element spacing of the radar array d is half of the radar wavelength. The geometry of the system is shown in Figure 1, where (θ, ϕ) is the beam direction, and v is the platform velocity. The polarimetric radar transmits uniformly M pulses. The pulse repetition frequency (PRF) in a coherent processing interval (CPI) is ι .

Figure 1. The geometry of polarimetric radar array.

The radar data vector under the target free hypothesis H_0 and the target presence hypothesis H_1 can be written by

$$\begin{cases} H_0 : \mathbf{x} = \mathbf{u} \\ H_1 : \mathbf{x} = \mathbf{t} + \mathbf{u} \end{cases} \quad (1)$$

where $\mathbf{u} = \mathbf{c} + \mathbf{n}$ is the clutter plus noise vector which is, as usual, a random vector drawn from a complex Gaussian process with zero-mean and unknown positive definite covariance matrix $\mathbf{R}$; $\mathbf{t} = \mathbf{A}\mathbf{p}$

is the target polarization-space-time steering vector (or the target vector), $\mathbf{p}$ is a complex random vector characterizing the unknown amplitude, phase, and polarization parameters of target return, and $\mathbf{A} = \mathbf{I}_2 \otimes \mathbf{d}_{f_D} \otimes \mathbf{s}_{f_S}$, $f_D = 2v \cos \theta \cos \varphi / (\lambda \iota)$ and $f_S = d \cos \theta \cos \varphi / \lambda$ denote the target normalized Doppler frequency and normalized spatial frequency, respectively; where $\otimes$ denotes the Kronecker product, $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, λ is the radar wavelength, and

$$\mathbf{d}_{f_D} = M^{-1/2} [1, e^{i2\pi f_D}, \dots, e^{i2\pi(M-1)f_D}]^T \quad (2)$$

$$\mathbf{s}_{f_S} = N^{-1/2} [1, e^{i2\pi f_S}, \dots, e^{i2\pi(N-1)f_S}]^T \quad (3)$$

with $\iota = \sqrt{-1}$, where superscript “T” denotes transpose.

Following Ward’s clutter model [5] and polarization characterize [1], the $2MN \times 1$ vector $\mathbf{c}$ can be written as follows

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_H \\ \mathbf{c}_V \end{bmatrix} = \sum_{n=1}^{N_C} \underbrace{\begin{bmatrix} p_{H,n} \\ p_{V,n} \end{bmatrix}}_{\mathbf{p}_n} \otimes \mathbf{d}_{f_{D,n}} \otimes \mathbf{s}_{f_{S,n}} \quad (4)$$

where N_C is the number of the independent clutter patches in one range cell, the $MN \times 1$ zero-mean complex Gaussian random vectors $\mathbf{c}_H$ and $\mathbf{c}_V$, given by

$$\mathbf{c}_H = \sum_{n=1}^{N_C} p_{H,n} (\mathbf{d}_{f_{D,n}} \otimes \mathbf{s}_{f_{S,n}}) \quad (5)$$

$$\mathbf{c}_V = \sum_{n=1}^{N_C} p_{V,n} (\mathbf{d}_{f_{D,n}} \otimes \mathbf{s}_{f_{S,n}}) \quad (6)$$

represent the horizontal and vertical clutter components whose covariance matrices are

$$\mathbf{R}_{HH} = E\{\mathbf{c}_H \mathbf{c}_H^H\} = \sum_{n=1}^{N_C} \underbrace{E\{|p_{H,n}|^2\}}_{\sigma_{H,n}^2} (\mathbf{d}_{f_{D,n}} \mathbf{d}_{f_{D,n}}^H) \otimes (\mathbf{s}_{f_{S,n}} \mathbf{s}_{f_{S,n}}^H) \quad (7)$$

$$\mathbf{R}_{VV} = E\{\mathbf{c}_V \mathbf{c}_V^H\} = \sum_{n=1}^{N_C} \underbrace{E\{|p_{V,n}|^2\}}_{\sigma_{V,n}^2} (\mathbf{d}_{f_{D,n}} \mathbf{d}_{f_{D,n}}^H) \otimes (\mathbf{s}_{f_{S,n}} \mathbf{s}_{f_{S,n}}^H) \quad (8)$$

in which $E\{\cdot\}$ denotes expectation and superscript “H” denotes conjugate transpose; $f_{D,n}$ and $f_{S,n}$ are, respectively, the Doppler frequency and the spatial frequency of the n th clutter patch,

$$\mathbf{d}_{f_{D,n}} = M^{-1/2} [1, e^{i2\pi f_{D,n}}, \dots, e^{i2\pi(M-1)f_{D,n}}]^T \quad (9)$$

$$\mathbf{s}_{f_{S,n}} = N^{-1/2} [1, e^{i2\pi f_{S,n}}, \dots, e^{i2\pi(N-1)f_{S,n}}]^T \quad (10)$$

The 2×1 vector $\mathbf{p}_n$ can be written as $\mathbf{p}_n = \mathbf{P}^{1/2} \boldsymbol{\rho}_n$, where $E\{\boldsymbol{\rho}_n \boldsymbol{\rho}_n^H\} = \sigma_{H,n}^2 \mathbf{I}_2$, and

$$\mathbf{P} = \begin{bmatrix} 1 & \sqrt{\gamma} \zeta \\ \sqrt{\gamma} \zeta^* & \gamma \end{bmatrix} \quad (11)$$

in which ζ and γ are the correlation coefficient and the power ratio between the two polarimetric channels.

3. Clutter Subspace Based pSTAP

The clutter covariance matrix is

$$\begin{aligned}\mathbf{R}_c &= E\{\mathbf{c}\mathbf{c}^H\} = \sum_{n=1}^{N_c} E\{\mathbf{p}_n \mathbf{p}_n^H\} \otimes \mathbf{d}_{f_D,n} \mathbf{d}_{f_D,n}^H \otimes \mathbf{s}_{f_S,n} \mathbf{s}_{f_S,n}^H \\ &= \mathbf{P} \otimes \left\{ \sum_{n=1}^{N_c} \sigma_{H,n}^2 (\mathbf{d}_{f_D,n} \mathbf{d}_{f_D,n}^H) \otimes (\mathbf{s}_{f_S,n} \mathbf{s}_{f_S,n}^H) \right\} = \mathbf{P} \otimes \mathbf{R}_{HH}\end{aligned}\quad (12)$$

According to [23], the effective rank of $\mathbf{R}_{HH}$ is $\lfloor N + \beta(M-1) \rfloor$, where $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer and $\beta = f_D/f_S = 2v/(dl)$ denotes the slope of the clutter ridge. From (12), the effective rank of $\mathbf{R}_c$ is

$$\text{rank}(\mathbf{R}_c) = \text{rank}(\mathbf{P}) \times \text{rank}(\mathbf{R}_{HH}) \quad (13)$$

Because the rank of $\mathbf{P}$ is 2, the effective rank of the clutter covariance matrix is

$$Q = \text{rank}(\mathbf{R}_c) = 2 \times \lfloor N + \beta(M-1) \rfloor \quad (14)$$

As the clutter and noise are independent, we have

$$\mathbf{R} = \mathbf{R}_c + \mathbf{R}_n = \sum_{i=1}^{2MN} \gamma_i \mathbf{v}_i \mathbf{v}_i^H = \mathbf{U}_c \mathbf{\Lambda}_c \mathbf{U}_c^H + \sigma^2 \mathbf{I}_{2MN} \quad (15)$$

Let $\mathbf{U}_c = [\mathbf{v}_1, \dots, \mathbf{v}_Q]$ whose column span is the so-called clutter subspace, $\mathbf{\Lambda}_c = \text{diag}(\gamma_1, \dots, \gamma_Q)$, which consists of the Q principal eigenvalues of $\mathbf{R}$, in which $\text{diag}(\cdot)$ denotes a diagonal matrix; and σ^2 is the variance of the Gaussian white noise.

The optimal pSTAP filter vector is given by [3],

$$\mathbf{w}_{\text{opt}} = \mathbf{R}^{-1} \mathbf{t} = \mathbf{R}^{-1} \mathbf{A} \mathbf{p} \quad (16)$$

Applying the matrix inversion lemma to (15) yields

$$\mathbf{R}^{-1} = (\mathbf{U}_c \mathbf{\Lambda}_c \mathbf{U}_c^H + \sigma^2 \mathbf{I}_{2MN})^{-1} = \sigma^{-2} [\mathbf{I}_{2MN} - \mathbf{U}_c (\mathbf{\Lambda}_c + \sigma^2 \mathbf{I}_{2MN})^{-1} \mathbf{\Lambda}_c \mathbf{U}_c^H] \quad (17)$$

Since the diagonal elements in $\mathbf{\Sigma}$ are much larger than σ^2 , (17) can be approximated as

$$\mathbf{R}^{-1} \approx \sigma^{-2} (\mathbf{I}_{2MN} - \mathbf{U}_c \mathbf{U}_c^H) \quad (18)$$

Therefore, the pSTAP filter vector can be written by

$$\mathbf{w} = (\mathbf{I}_{2MN} - \mathbf{U}_c \mathbf{U}_c^H) \mathbf{A} \mathbf{p} \quad (19)$$

In the following, we give a new method for estimating $\mathbf{U}_c$ and $\mathbf{p}$.

4. The Random Nystrom Method

4.1. Revisit of Standard Nystrom

The CNCM $\mathbf{R}$ can be approximated as follows

$$\mathbf{R} \approx \underbrace{\begin{bmatrix} \mathbf{U}_{11} \\ \mathbf{U}_{21} \end{bmatrix}}_{\mathbf{U}} \underbrace{\mathbf{\Lambda}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \mathbf{U}_{11} \\ \mathbf{U}_{21} \end{bmatrix}}^H = \hat{\mathbf{R}} = \begin{bmatrix} \underbrace{\mathbf{U}_{11} \mathbf{\Lambda} \mathbf{U}_{11}^H}_{\hat{\mathbf{R}}_{11}} & \underbrace{\mathbf{U}_{11} \mathbf{\Lambda} \mathbf{U}_{21}^H}_{\hat{\mathbf{R}}_{21}^H} \\ \underbrace{\mathbf{U}_{21} \mathbf{\Lambda} \mathbf{U}_{11}^H}_{\hat{\mathbf{R}}_{21}} & \underbrace{\mathbf{U}_{21} \mathbf{\Lambda} \mathbf{U}_{21}^H}_{\hat{\mathbf{R}}_{22} = \hat{\mathbf{R}}_{21} \hat{\mathbf{R}}_{11}^{-1} \hat{\mathbf{R}}_{21}^H} \end{bmatrix} \quad (20)$$

where $\mathbf{U}$ is an $2MN \times Q$ orthogonal matrix whose columns are the Q principal eigenvectors of $\mathbf{R}$, $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the Q principal eigenvalues of $\mathbf{R}$; $\mathbf{U}_{11} \in \mathbb{C}^{K \times Q}$, and $\mathbf{U}_{21} \in \mathbb{C}^{(2MN-K) \times Q}$, with $Q \leq K < 2MN$.

Let

$$\hat{\mathbf{C}} = \hat{\mathbf{R}} \underbrace{\begin{bmatrix} \mathbf{S}_{11} \\ \mathbf{S}_{21} \end{bmatrix}}_{\mathbf{S}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} \\ \hat{\mathbf{C}}_{21} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{R}}_{11} \mathbf{S}_{11} + \hat{\mathbf{R}}_{21}^H \mathbf{S}_{21} \\ \hat{\mathbf{R}}_{21} \mathbf{S}_{11} + \hat{\mathbf{R}}_{21} \hat{\mathbf{R}}_{11}^{-1} \hat{\mathbf{R}}_{21}^H \mathbf{S}_{21} \end{bmatrix} \quad (21)$$

where $\mathbf{S}$ is a $2MN \times K$ selection matrix, and $\mathbf{S}_{11} \in \mathbb{C}^{K \times K}$, and $\mathbf{S}_{21} \in \mathbb{C}^{(2MN-K) \times K}$. In the standard Nyström, $\mathbf{S}$ is composed of K different but arbitrary columns of $\mathbf{I}_{2MN}$.

Define then

$$\hat{\mathbf{U}} = \begin{bmatrix} \hat{\mathbf{U}}_{11} \\ \hat{\mathbf{U}}_{21} \end{bmatrix} \quad (22)$$

where $\hat{\mathbf{U}}_{21} = \hat{\mathbf{C}}_{21} \hat{\mathbf{U}}_{11} \hat{\mathbf{\Lambda}}_{11}^{-1}$, in which $\hat{\mathbf{U}}_{11}$ is an $K \times Q$ orthogonal matrix whose columns are the eigenvectors of $\hat{\mathbf{C}}_{11}$, and $\hat{\mathbf{\Lambda}}_{11}$ is the $Q \times Q$ diagonal matrix whose diagonal elements are the eigenvalues of $\hat{\mathbf{C}}_{11}$.

It can be verified that $\text{rank}(\hat{\mathbf{U}}) = Q$, and

$$\hat{\mathbf{U}} \hat{\mathbf{\Lambda}}_{11} = \begin{bmatrix} \hat{\mathbf{U}}_{11} \hat{\mathbf{\Lambda}}_{11} \\ \hat{\mathbf{U}}_{21} \hat{\mathbf{\Lambda}}_{11} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{C}}_{11} \\ \hat{\mathbf{C}}_{21} \end{bmatrix} \hat{\mathbf{U}}_{11} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{S} \hat{\mathbf{U}}_{11} \quad (23)$$

It follows that

$$\hat{\mathbf{U}} = \mathbf{U} (\mathbf{\Lambda} \mathbf{U}^H \mathbf{S} \hat{\mathbf{U}}_{11} \hat{\mathbf{\Lambda}}_{11}^{-1}) \quad (24)$$

This implies that $\text{span}(\hat{\mathbf{U}}) = \text{span}(\mathbf{U})$, which is exactly the Q -dimensional clutter subspace to be identified.

The computational complexity of the standard Nyström is about $O(2MNK^2)$. If $K \ll 2MN$, it is computationally much simpler than the classical eigendecomposition method whose computational complexity is about $O((2MN)^3)$. However, making use of a relatively small K in the method would result in poor identification $\text{span}(\mathbf{U})$ since only K columns of $\mathbf{R}$ are employed. In the sequel, a R-Nyström method is developed which fulfills a good compromise between the computational complexity and identification performance. In the new method, the dimension of the selection matrix $\mathbf{S}$ is still of $2MN \times K$. However, each column of $\mathbf{S}$ is designed by a new criterion based on $\mathbf{R}$. Note that all Nyström-based methods require EVD operation, except that R-Nyström is based on a low-dimensional matrix for EVD.

4.2. The Random Nyström Method

To show the estimated performance of the clutter subspace from $\mathbf{R}$, a cost function (projection error) which denotes the difference between the estimated and the real projection operator of the clutter subspace is given. The smaller the function, the more accurate the estimation. The cost function is given by

$$f(\mathbf{U}_e) = \min_{\mathbf{U}_e} \left\| \mathbf{U}_e (\mathbf{U}_e^H \mathbf{U}_e)^{-1} \mathbf{U}_e^H - \mathbf{U}_r \mathbf{U}_r^H \right\|_F \quad (25)$$

where $\mathbf{U}_e$ and $\mathbf{U}_r$ denote the estimated and real matrix representation of the clutter subspace, respectively; $\|\cdot\|_F$ denotes the Frobenius norm. Because the estimated matrix $\mathbf{U}_e$ is not a column orthogonal matrix, the projection operator is used by $\mathbf{U}_e (\mathbf{U}_e^H \mathbf{U}_e)^{-1} \mathbf{U}_e^H$ in (25). Observed from the standard Nyström method in part 4.1, $\mathbf{U}_e$ is determined by the selection matrix $\mathbf{S}$. Therefore, $f(\mathbf{U}_e)$ is also determined by $\mathbf{S}$. In the following, a new criterion is proposed to determine the selection matrix $\mathbf{S}$.

Define a $2MN \times K$ matrix $\mathbf{S}$, and assume all the elements in $\mathbf{S}$ are determined with $\mathbf{B}(i, j)$ as the probability distribution function, namely the value at (i, j) depends on $\mathbf{B}(i, j)$. Let $\mathbf{S}(i, j)$ be the function at (i, j) of matrix $\mathbf{S}$ whose definition is represented as follows [24]

$$\mathbf{S}(i, j) = \begin{cases} 1, & \mathbf{E}(i, j) \leq \mathbf{B}(i, j) \\ 0, & \mathbf{E}(i, j) > \mathbf{B}(i, j) \end{cases} \quad (26)$$

where $\mathbf{E}(i, j)$ is a uniformly distributed random number between [0,1].

The derivation of $\mathbf{B}(i, j)$ is given in the following. First, perform singular value decomposition (SVD) for $\mathbf{R}$,

$$\mathbf{R} = \mathbf{U}' \mathbf{\Sigma}' \mathbf{V}'^H \quad (27)$$

where $\mathbf{U}' = [\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_{2MN}]$, $\mathbf{V}' = [\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_{2MN}]$, and $\mathbf{\Sigma}' = \text{diag}\{\sigma'_1, \sigma'_2, \dots, \sigma'_{2MN}\}$ with $\sigma'_1 \geq \sigma'_2 \geq \dots \geq \sigma'_{2MN} \geq 0$.

The j th column of $\mathbf{R}$ can be represented as follows

$$\mathbf{R}_j = \sum_{q=1}^Q \sigma'_q \mathbf{u}'_q (\mathbf{v}'_{q,j})^* = \sum_{q=1}^Q [\sigma'_q (\mathbf{v}'_{q,j})^*] \mathbf{u}'_q \quad (28)$$

where $\mathbf{v}'_{q,j}$ is the j th element of the q th right singular vector, and “*” denotes a conjugate. Observed from (28), $\mathbf{R}_j$ is a linear combination of the left singular vectors and right singular values, and the coefficients are the corresponding j th row of $\mathbf{V}'$. Hence, $\mathbf{R}_j$ can be approximated with a linear combination of the larger K left singular vectors and its corresponding singular values

$$\mathbf{R}_j \approx \sum_{q=1}^K [\sigma'_q (\mathbf{v}'_{q,j})^*] \mathbf{u}'_q \quad (29)$$

All the coefficients are extracted and used to construct the normalized statistical leverage score, which is represented as follows

$$\varepsilon_j = \frac{1}{K} \sum_{k=1}^K [(\mathbf{v}'_{k,j})^*]^2 \quad (30)$$

where $\varepsilon_j > 0$, $\sum_{j=1}^{2MN} \varepsilon_j = 1$ and $1 \leq j \leq 2MN$. The whole leverage scores form a probability distribution over the $2MN$ columns [18]. A new RRCLS based on related works on RLS [19–22] can be defined by

$$\tau_{i,j} = \{[\mathbf{D}\mathbf{R}(\mathbf{R} + \kappa\mathbf{I})^{-1}]^T\}_{i,j} \quad (31)$$

in which $\mathbf{D}$ is a $K \times 2MN$ matrix whose elements follow a standard normal distribution, κ is a constant and $1 \leq i \leq 2MN$, $1 \leq j \leq K$. $\mathbf{B}(i, j)$ can be obtained by

$$\mathbf{B}(i, j) = \min\{1, \delta\tau_{i,j}\} \quad (32)$$

where δ is an adjustable factor. After obtaining the probability distribution function $\mathbf{B}(i, j)$, the selection matrix $\mathbf{S}$ can be determined by (26). Compared to the standard Nyström, the number of non-zero elements in each column of $\mathbf{S}$ is more than one.

The selected low dimension matrix can be represented as

$$\tilde{\mathbf{R}} = \mathbf{R}\mathbf{S} = \begin{bmatrix} \tilde{\mathbf{R}}_{11} \\ \tilde{\mathbf{R}}_{21} \end{bmatrix} \quad (33)$$

Similar to (22), define

$$\tilde{\mathbf{U}} = \begin{bmatrix} \tilde{\mathbf{U}}_{11} \\ \tilde{\mathbf{U}}_{21} \end{bmatrix} \quad (34)$$

where $\tilde{\mathbf{U}}_{21} = \tilde{\mathbf{R}}_{21} \tilde{\mathbf{U}}_{11}^{-1}$, in which $\tilde{\mathbf{U}}_{11}$ is an $K \times Q$ orthonormal matrix whose columns are the eigenvectors of $\tilde{\mathbf{R}}_{11}$, and $\tilde{\mathbf{\Lambda}}_{11}$ is the $Q \times Q$ diagonal matrix whose diagonal elements are the eigenvalues of $\tilde{\mathbf{R}}_{11}$. Note that all the columns of $\tilde{\mathbf{U}}$ are not orthogonal to each other.

The R-Nyström algorithm for the matrix representation of the clutter subspace is summarized as follows:

- determine the selection matrix $\mathbf{S}$ by (26), (31) and (32)
- obtain the low-dimensional matrix by (33)
- estimate the matrix representation of the clutter subspace by (34)

4.3. Application of R-Nyström to pSTAP

The CNCM $\mathbf{R}$ is given by the sample estimate

$$\bar{\mathbf{R}} = E\{\mathbf{u}\mathbf{u}^H\} \approx \frac{1}{L} \sum_{l=1}^L \mathbf{u}_l \mathbf{u}_l^H \quad (35)$$

where $\mathbf{u}_l$ is the clutter plus noise vector of the l th range cell and L is the number of training samples.

Substitute (35) into (31), the new RRCLS can be written as

$$\bar{\tau}_{i,j} = \{[\mathbf{D}\bar{\mathbf{R}}(\bar{\mathbf{R}} + \kappa\mathbf{I})^{-1}]^T\}_{i,j} \quad (36)$$

By (36), the corresponding $\bar{\mathbf{B}}(i, j)$ can be written as

$$\bar{\mathbf{B}}(i, j) = \min\{1, \delta\bar{\tau}_{i,j}\} \quad (37)$$

Therefore, the new selection matrix $\bar{\mathbf{S}}$ is calculated by

$$\bar{\mathbf{S}}(i, j) = \begin{cases} 1, & \mathbf{E}(i, j) \leq \bar{\mathbf{B}}(i, j) \\ 0, & \mathbf{E}(i, j) > \bar{\mathbf{B}}(i, j) \end{cases} \quad (38)$$

Because there is an error in the sample estimate, $\mathbf{x}$ in H_0 is processed directly for avoiding estimation error into the following calculation. The selected output is written by

$$\mathbf{y} = \bar{\mathbf{S}}^H \mathbf{x} \quad (39)$$

and the corresponding low-dimensional matrices are given by

$$\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\} \approx \bar{\mathbf{S}}^H \bar{\mathbf{R}} \bar{\mathbf{S}} \quad (40)$$

$$\mathbf{R}_{xy} = E\{\mathbf{x}\mathbf{y}^H\} = \bar{\mathbf{R}} \bar{\mathbf{S}} \quad (41)$$

Then, performing EVD of $\mathbf{R}_{yy}$ yields,

$$\mathbf{R}_{yy} = \sum_{i=1}^K \lambda_i \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^H \quad (42)$$

where λ_i ($\lambda_1 \geq \dots \geq \lambda_Q > \lambda_{Q+1} = \dots = \lambda_K$) is the eigenvalue of $\mathbf{R}_{yy}$ and $\bar{\mathbf{u}}_i$ is the corresponding eigenvector. Let $\bar{\mathbf{U}}_{yy} = [\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \dots, \bar{\mathbf{u}}_Q]$ and $\bar{\Lambda}_{yy} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_Q)$, and the matrix representation of the clutter subspace is given by

$$\bar{\mathbf{U}} = \mathbf{R}_{xy} \bar{\mathbf{U}}_{yy} \bar{\Lambda}_{yy}^{-1} \quad (43)$$

The columns of $\bar{\mathbf{U}}$ are also not mutually orthogonal. Based on (19), the weight vector of the pR-Nyström method is given by

$$\mathbf{w}_{\text{pR-Nyström}} = (\mathbf{I}_{2MN} - \bar{\mathbf{U}}(\bar{\mathbf{U}}^H \bar{\mathbf{U}})^{-1} \bar{\mathbf{U}}^H) \mathbf{A} \mathbf{p} \quad (44)$$

It notes that the polarimetric signal vector to be detected is unknown in advance, hence, $\mathbf{p}$ needs to be estimated by the proposed modified least square method (mLSM). In hypothesis H_1 , the radar data are represented as follows

$$\mathbf{x} = \mathbf{A}\mathbf{p} + \mathbf{c} + \mathbf{n} \quad (45)$$

Introduce function $f(\mathbf{p})$, which is written by

$$f(\mathbf{p}) = \|\mathbf{A}\mathbf{p} - \mathbf{x}'\|_2^2 \quad (46)$$

where $\mathbf{x}' = \mathbf{x} - \mathbf{c}$ denotes the data that contain no clutter components. For eliminating the clutter effect in $\mathbf{x}$, $\mathbf{x}'$ can be obtained by projecting $\mathbf{x}$ into the orthogonal subspace of clutter subspace, where the matrix representation of the clutter subspace $\hat{\mathbf{U}}$ can be obtained by the standard Nyström method,

$$\mathbf{x}' = [\mathbf{I}_{2MN} - \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H]\mathbf{x} \quad (47)$$

Substitute (47) into (46),

$$f(\mathbf{p}) = \|\mathbf{A}\mathbf{p} - [\mathbf{I}_{2MN} - \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H]\mathbf{x}\|_2^2 \quad (48)$$

Therefore, when $\hat{\mathbf{p}} = \mathbf{p}$, $f(\mathbf{p})$ can get the minimum value. Let

$$\hat{\mathbf{p}} = \arg \min\{\|\mathbf{A}\mathbf{p} - [\mathbf{I}_{2MN} - \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H]\mathbf{x}\|_2^2\} \quad (49)$$

Taking the sub-gradient of (49) with respect to $\mathbf{p}$, $\hat{\mathbf{p}}$ can be obtained as follows

$$\hat{\mathbf{p}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H[\mathbf{I}_{2MN} - \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H]\mathbf{x} \quad (50)$$

The final filter vector is rewritten as

$$\hat{\mathbf{w}}_{\text{pR-Nyström}} = (\mathbf{I}_{2MN} - \bar{\mathbf{U}}(\bar{\mathbf{U}}^H\bar{\mathbf{U}})^{-1}\bar{\mathbf{U}}^H)\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H[\mathbf{I}_{2MN} - \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H]\mathbf{x} \quad (51)$$

Extend the existing methods to the pSTAP case, which are called pEVD, pPAST, pFAPI, pL1-PAST, and pBiPerEFA, respectively. Note that all the methods use the estimated $\hat{\mathbf{p}}$ in the process of calculating the corresponding filter.

5. Computational Complexity Analysis

The major computational burden of the pR-Nyström method is engaged with the calculation for the selection matrix and the polarimetric signal vector to be detected. The main polarimetric radar system parameters are as shown in Table 1, and are the same as that in [2]. The approximate computational complexity and the averaged computation time of 100 independent runs are listed in Table 2, where the number of training samples is $L = 130$ and $K = Q$ under the following configurations: (1) CPU: Intel(R) Core (TM) i7-6700 3.40 GHz; (2) memory: 8.00 GB; (3) system: 64-bit Windows 7; (4) MATLAB R2016a.

Table 1. Polarimetric radar system parameters.

Parameter	Value
antenna array	side-looking array
the number of antennas	16
pulse number in one CPI	16
CNR	40 dB
PRF	1200
velocity	200 m/s
airborne height	9 km
looking direction	0°
target auxiliary polarization angle	30°
target polarization phase difference	60°
clutter degree of polarization	0.99
the power ratio between two channels	1
clutter statistical average phase difference	90°

Table 2. Complexity and computation time comparison.

Method	Complexity	Time/ms
mLSM	$t_1 = O\left(\begin{array}{l} K^2(3K+2L-2) + MNK(4L+8MN+16K-8) \\ + 4MN(4M^2N^2 + MN + 4) + 16 \end{array} \right)$	15.0
pEVD	$O\left(\begin{array}{l} 8M^3N^3 + 16M^2N^2L + 8M^2N^2K \\ + 56M^2N^2 - 2MN(L-12) + 4 \end{array} \right) + t_1$	59.5
pPAST	$O\left(\begin{array}{l} 60M^2N^2 + (8MN + 6K^2 - 3)KL \\ + 8M^2N^2(K+L) - 4MN(L-6) + 4 \end{array} \right) + t_1$	47.8
pFAPI	$O\left(\begin{array}{l} (8K^2 + 8K - 7)K + 8M^2N^2(2K+L) \\ + 2MN(K-L+12) + 60M^2N^2 + 4 \end{array} \right) + t_1$	50.1
pL1-PAST	$O\left(\begin{array}{l} 60M^2N^2 + (6K^2 + 6MN + 4MNK - 2K - 2)KL \\ - 4MN(L-6) + 8M^2N^2(K+L) \end{array} \right) + t_1$	126.9
pNyström	$O\left(\begin{array}{l} 4K^3 + (2K-1)(10MNK + 8M^2N^2) \\ + (2L-1)(K^2 + 2MNK) + 2MN(4MN-1) \end{array} \right) + t_1$	21.5
pR-Nyström	$O\left(\begin{array}{l} 4K^3 + K(2KL - K - L) + 2MN(16KL - 6K - L + 12) \\ + 4M^2N^2(2MN + 2K + 6L + 14) + 4 \end{array} \right) + t_1$	36.4
pBiPerEFA	$O\left(\begin{array}{l} 8MNL(MN+3) + 2MN(22MN + 84N + 11) \\ - 6N(6N+L) + 54(108N^3 + 126N^2 + 26N - 3) \end{array} \right) + t_1$	29.4

It can be seen from Table 2 that the average computation time of the mLSM method is denoted as t_1 , and the average computation time of the pR-Nyström method is less than that of the other existing methods except for the pNyström method and pBi-PerEFA method. The EFA referred to in this paper is with three Doppler bins and the iteration number of pBiPerEFA method [25] is nine.

6. Numerical Results

In this section, to assess SCNR loss, simulated data are employed as the metric for comparison among different methods, which is defined by

$$\text{SCNR loss} = \frac{\text{SCNR}_o}{\text{SCNR}_i} = \frac{|\hat{\mathbf{w}}^H \hat{\mathbf{t}}|^2}{(\hat{\mathbf{w}}^H \mathbf{R}_{\text{real}} \hat{\mathbf{w}})(\hat{\mathbf{t}}^H \hat{\mathbf{t}})} \quad (52)$$

where SCNR_o and SCNR_i denote the output and input signal to clutter plus noise ratio, respectively; $\mathbf{R}_{\text{real}}$ represents the true CNCM which is assumed to be known in simulation experiments and $\hat{\mathbf{t}} = \mathbf{A}\hat{\mathbf{p}}$. The polarimetric radar system parameters are shown in Table 1, and $\kappa = 1$. The curves shown are the averaged results of 100 Monte Carlo trials.

Shown in Figure 2 are the projection error curves versus training samples with different values of δ , where the target normalized Doppler frequency is 0.2. It is observed that the pR-Nyström method is robust for different values of δ between [10, 40]. In the following simulation, we set $\delta = 20$.

Shown in Figure 3 are the projection error curves versus training samples, where the target normalized Doppler frequency is 0.2. It is observed that the pR-Nyström method has a smaller projection error and is closest to the pEVD method. This indicates that the pR-Nyström method has a better approximate performance than the other methods. Furthermore, in the case of the non-side-looking model, assuming that the angle between the array pointing and the direction of movement is 30° , the SCNR loss is a little weaker than the side-looking model.

Shown in Figure 4 are the SCNR loss curves versus training samples where the target normalized Doppler frequency is 0.2. It is observed that the pR-Nyström method has a better SCNR loss performance than the other methods with the increase in the training sample and is closest to the pEVD method.

Figure 2. Projection error versus training samples: with different δ .

Figure 3. Projection error versus training samples.

Figure 4. Signal to clutter plus noise ratio (SCNR) loss performance versus training samples.

Shown in Figure 5 are the SCNR loss curves versus the target normalized Doppler frequency, where the number of training samples is 130. It is observed that the pR-Nyström method can achieve a better SCNR loss performance than the other methods with different target normalized Doppler frequencies and is closest to the pEVD method.

Figure 5. SCNR loss performance versus target Doppler frequency.

Shown in Figure 6 are the SCNR loss curves versus the normalized Doppler frequency, with different clutter degree of polarization (DOP). The number of training samples is 130, and $\text{CNR} = 40$ dB. It is observed that when the target Doppler frequency is large, the SCNR loss performance of the pR-Nyström method stays stable with different DOPs; when the normalized Doppler frequency value is located in $[-0.1, 0.1]$, the SCNR loss performance of pR-Nyström method changes. When the clutter DOP is large, the clutter is a completely polarized wave and this will make a larger notch [2]. When the clutter DOP is gradually reduced, the clutter presents a partially polarized wave and this leads to deeper and deeper notch.

Figure 6. SCNR loss performance versus target Doppler frequency: with different clutter degrees of polarization (DOPs).

Shown in Figure 7 are the SCNR loss curves versus the normalized Doppler frequency, with different CNRs. The number of training samples is 130, and $|\zeta| = 0.99$. It is observed that the pR-Nyström method has a stable SCNR loss performance when the normalized Doppler frequency value is large.

When the normalized Doppler frequency value is located in $[-0.1, 0.1]$, the SCNR loss performance of the pR-Nyström method changes with different CNRs. The reason is that when the clutter notch is formed in the main lobe clutter region of Doppler frequency by the clutter suppressor, its depth depends on the CNR. The clutter in the side lobe clutter region can be well suppressed to the noise level and is less affected by the CNR.

Figure 7. SCNR loss performance versus target Doppler frequency: with different clutter to noise ratio (CNR).

7. Conclusions

We have presented a random Nyström framework based on the matrix representation of the clutter subspace estimation method for pSTAP. The new method involves a new selection matrix for the complete use of CNCM, which is determined by the proposed random ridge cross leverage score rule. Simulation results show that the proposed method has an appealing clutter suppression performance and computational efficiency as compared with the existing methods.

It can be seen from the investigation in this article that high subspace estimation accuracy and computational efficiency can be achieved based on the new column selection criterion within the Nyström framework. In order to reduce the influence of adjustable factors on subspace estimation, the number of adjustable factors in the column selection criterion should be as small as possible. Hence, it is important to investigate further the adjustable factors in the subsequent research. When the newly proposed method is applied in airborne radar, the clutter suppression performance of pSTAP can be improved, and the target detection probability can be increased.

Author Contributions: Conceptualization, K.Z. and Y.X.; methodology, K.Z.; software, K.Z.; formal analysis, K.Z.; investigation, K.Z., Z.L., S.S., Y.H., and Y.X.; writing—original draft preparation, K.Z.; writing—review and editing, Y.X.; supervision, Y.X. and Z.L.; project administration, Z.L.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 61490691.

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

References

1. Park, H.R.; Wang, H.; Li, J. An adaptive polarization-space-time processor for radar system. In Proceedings of the IEEE Antennas and Propagation Society International Symposium, Ann Arbor, MI, USA, 28 June–2 July 1993; pp. 698–701.
2. Wu, D.; Xu, Z.; Zhang, L.; Xiong, Z.; Xiao, S. Performance analysis of polarization-space-time three-domain joint processing for clutter suppression in air-borne radar. *Prog. Electromagn. Res.* **2012**, *129*, 579–601. [[CrossRef](#)]

3. Wu, D.; Xu, Z.; Xiong, Z.; Zhang, L.; Xiao, S. Polarization-space-time joint domain processing for clutter suppression in subspace leakage environments. In Proceedings of the 2011 IEEE CIE International Conference on Radar, Chengdu, China, 24–27 October 2011; pp. 692–695.
4. Du, W.; Liao, G.; Yang, Z. Clutter space-time covariance matrix estimate based on multi-polarised data. *IET Radar Sonar Navig.* **2014**, *8*, 1109–1115. [[CrossRef](#)]
5. Ward, J. *Space-Time Adaptive Processing for Airborne Radar*; IET: London, UK, 1998.
6. Melvin, W.L. Space-time adaptive radar performance in heterogeneous clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 621–633. [[CrossRef](#)]
7. Wang, Z.; Wang, Y.; Duan, K.; Xie, W. Subspace-augmented clutter suppression technique for STAP radar. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 462–466. [[CrossRef](#)]
8. Yang, B. Projection approximation subspace tracking. *IEEE Trans. Signal Process.* **1995**, *43*, 95–107. [[CrossRef](#)]
9. Badeau, R.; David, B.; Richard, G. Fast approximated power iteration subspace tracking. *IEEE Trans. Signal Process.* **2005**, *53*, 2931–2941. [[CrossRef](#)]
10. Yang, X.; Sun, Y.; Zeng, T.; Long, T.; Sarkar, T.K. Fast STAP method based on PAST with sparse constraint for airborne phased array radar. *IEEE Trans. Signal Process.* **2016**, *64*, 4550–4561. [[CrossRef](#)]
11. Zhang, K.; Kwok, J.T. Clustered Nyström method for large scale manifold learning and dimension reduction. *IEEE Trans. Neural Netw.* **2010**, *21*, 1576–1587. [[CrossRef](#)] [[PubMed](#)]
12. Kumar, S.; Mohri, M.; Talwalkar, A. Sampling methods for the Nyström method. *J. Mach. Learn. Res.* **2012**, *13*, 981–1006.
13. Wang, S.; Zhang, Z. Improving CUR matrix decomposition and the Nyström approximation via adaptive sampling. *J. Mach. Learn. Res.* **2013**, *14*, 2729–2769.
14. Qian, C.; Huang, L.; So, H.C. Computationally efficient ESPRIT algorithm for direction-of-arrival estimation based on Nyström method. *Signal Process.* **2014**, *94*, 74–80. [[CrossRef](#)]
15. Baker, C.T.H. *The Numerical Treatment of Integral Equations*; Oxford University Press: Oxford, UK, 1978.
16. Williams, C.K.I.; Seeger, M. Using the Nyström method to speed up kernel machines. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2000.
17. Fowlkes, C.; Belongie, S.; Chung, F.; Malik, J. Spectral grouping using the Nyström method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 214–225. [[CrossRef](#)] [[PubMed](#)]
18. Drineas, P.; Magdon-Ismail, M.; Mahoney, M.W.; Woodruff, D.P. Fast approximation of matrix coherence and statistical leverage. *J. Mach. Learn. Res.* **2012**, *13*, 3475–3506.
19. Cohen, M.B.; Musco, C.; Musco, C. Ridge leverage scores for low-rank approximation. *arXiv* **2015**, arXiv:1511.07263.
20. Calandriello, D.; Lazaric, A.; Valko, M. Analysis of Nyström method with sequential ridge leverage score sampling. In Proceedings of the 2016 Thirty-Second Conference on Uncertainty in Artificial Intelligence, New York, NY, USA, 25–29 June 2016.
21. Musco, C.; Musco, C. Recursive sampling for the Nyström method. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
22. Wang, W.; Chen, Z.; Li, X.; Wang, B. Space time adaptive processing algorithm for multiple-input-multiple-output radar based on Nyström method. *IET Radar Sonar Navig.* **2016**, *10*, 459–467. [[CrossRef](#)]
23. Reed, I.S.; Mallet, J.D.; Brennan, L.E. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *AES-10*, 853–863. [[CrossRef](#)]
24. Smith, M.S.; Tan, T.G. Sidelobe reduction using random methods for antenna arrays. *Electron. Lett.* **1983**, *19*, 931–933. [[CrossRef](#)]
25. Zhou, Y.; Chen, X.; Li, Y.; Li, Y.; Wang, L.; Jiang, B.; Fang, D. A fast STAP method using persymmetry covariance matrix estimation for clutter suppression in airborne MIMO radar. *EURASIP J. Adv. Signal Process.* **2019**, *2019*, 13. [[CrossRef](#)]