Compressive Radar Clutter Subspace Estimation Using Dictionary Learning

Linda Bai, Sumit Roy
Department of Electrical Engineering
University of Washington
Seattle, WA, USA

Muralidhar Rangaswamy
AFRL/RYAP
Wright Patterson Air Force Base
Dayton, OH, USA

Abstract—Space-Time Adaptive Processing (STAP) based on matched filter processing in the presence of additive clutter (modeled as colored noise) requires knowledge of the clutter covariance matrix. In practice, this is estimated via the sample covariance matrix using samples from the neighboring range bins around the reference bin. By applying compressive sensing, the number of training samples needed to estimate the covariance matrix can be significantly reduced, provided that the basis mismatch problem, inherent to compressive sensing can be mitigated. This paper presents an adaptive approach to choosing the best sparsifying basis, using dictionary learning to estimate the radar clutter subspace. Numerical results show that the proposed algorithm achieves the desired reduction in training samples, and is more accurate than previous reduced-rank algorithm baseline.

Index Terms—Compressive Sensing, Dictionary Learning, Space-Time Adaptive Processing

I. INTRODUCTION

Space-Time Adaptive Processing (STAP) is a widely used radar signal processing technique to detect a multi-channel signal of interest (target) in temporally and spatially correlated clutter [1]. Due to motion of the aerial platform, the target cannot be efficiently detected from the clutter background by using spatial or temporal information alone, leading to the necessity of space-time joint processing. A matched filter to suppress clutter using its statistics is fundamental to effective STAP. To construct the matched filter, an estimate of the clutter covariance matrix is needed, that usually requires numerous homogenous samples from neighboring range cells. This is generally infeasible, consequently there is a rich literature on reducing the number of samples by exploiting prior information [2][3]. Many of these techniques involves exploiting the low-rank nature of the covariance matrix, such as [4] and [5].

Compressive sensing (CS) is a well-known technique that has rapidly risen to prominence in the signal processing communities for its ability to provide innovative solutions to problems of signal estimation, where the signal of interest is inherently sparse in some basis. For a vector signal in STAP, this is equivalent to exploiting the low-rank nature of the clutter subspace in the angular and (temporal) Doppler frequency domains that has the potential to significantly reduce the required number of samples (i.e., neighboring range bins) required for estimation of the clutter covariance matrix [6], [7]. However, in [6], the angle and frequency axis are discretized on a uniform grid, via the use of steering vectors corresponding to the lattice points as the sparsifying basis. As clutter patches may not be precisely located at the grid points, the resulting basis mismatch can lead to inaccurate clutter subspace estimation [8], in turn leading to significant degradation in STAP performance.

In this paper, we propose the use of dictionary learning algorithm to estimate the clutter subspace, where the angle and frequency axis are discretized non-uniformly and adaptively to mitigate the basis mismatch and obtain more accurate estimates of the clutter covariance matrix. The proposed compressive sensing with dictionary learning (CSDL) algorithm is compared with Multi-stage Wiener filter (MWF) [4], Conjugate Gradient Parametric Adaptive Matched Filter (CGPAMF) [5], Principal Component Inverse (PCI) [9], and Recursive Gram-Schmidt orthonormal basis selection algorithm (RGS). Numerical results based on the Knowledge Aided Sensor Signal Processing and Expert Reasoning (KASSPER) dataset [10] show that CSDL is more accurate than other algorithms with low sample support.

II. SYSTEM DESCRIPTION

A. System model

Fig. 1 shows a typical airborne radar antenna array. Consider a radar consisting of $N$ antenna array elements, each transmitting $M$ coherent pulses at a constant pulse repetition frequency (PRF) in a set range of directions. As given in [2], in a target-free scenario, the received signal is

$$\mathbf{x}_{MN \times 1} = \sum_{i=1}^{N_c} \gamma_i \phi(\theta_i, f_i) + \mathbf{n}_{MN \times 1} = \Phi \gamma + \mathbf{n}$$

(1)

where

$$\Phi = [\phi(\theta_1, f_1) \phi(\theta_2, f_2) \ldots \phi(\theta_{N_c}, f_{N_c})]$$

(2)

and

$$\gamma = [\gamma_1 \gamma_2 \ldots \gamma_{N_c}]^T.$$ 

(3)

$\mathbf{n} \sim CN(0, \sigma^2 \mathbf{I})$ is the additive white Gaussian noise, $N_c$ is the number of clutter patches, $\theta_i$ and $f_i$ are the complement of the cone angle and the Doppler frequency of the $i^{th}$ clutter
patch, $\gamma_i$ is the complex scattering intensity (i.e., coefficients with phase) of the $i^{th}$ clutter patch, and

$$\phi(\theta_i, f_i) = \left[ e^{j2\pi\frac{\lambda}{d} \sin \theta_i} \ldots e^{j2\pi(M-1)\frac{\lambda}{d} \sin \theta_i} \right]^T \otimes \left[ e^{j2\pi\frac{\lambda}{d} \frac{2}{\pi} \sin \theta_i} \ldots e^{j2\pi(N-1)\frac{\lambda}{d} \frac{2}{\pi} \sin \theta_i} \right]^T$$  

(4)

is the steering vector for a given angle and frequency, $\otimes$ is the Kronecker product, $d$ is the distance between two array elements, $\lambda$ is the radar wavelength.

The optimal STAP matched filter to suppress clutter is then given as in [2]

$$w = R^{-1}s$$  

(5)

where $R = E\{xx^H\}$ is the clutter covariance matrix, and $s$ is the steering vector for the target. Typically, $R$ is estimated using the sample matrix inversion (SMI) algorithm with $L$ target-free snapshots, which is given by

$$R_{SMI} = \frac{1}{L} \sum_{i=1}^{L} x_i x_i^H.$$  

(6)

Define the normalized Signal to Interference plus Noise Ratio (SINR) as

$$SINR = \frac{|s^H \tilde{R}^{-1} s|^2}{|s^H \tilde{R}^{-1} R \tilde{R}^{-1} s|^2 |s^H \tilde{R}^{-1} s|^2}$$  

(7)

where $\tilde{R}$ is the estimated clutter covariance matrix. The normalized SINR shows the degradation of performance due to inaccurate estimate of the clutter covariance matrix compared to knowing the true $R$. According to the Reed-Mallett-Brennan (RMB) rule [11], a normalized SINR above $-3$ dB using SMI requires $L \geq 2MN$ snapshots, and such a large number of homogeneous training samples are difficult to obtain in practice. On the other hand, by Brennan’s rule [9], $R$ is a low-rank matrix with rank $r \ll MN$ in most scenarios, leading to reduced-rank processing algorithms, such as PCI, RGS, MWF, CGPAMF, and the proposed CSDL algorithm, which use less snapshots than SMI.

B. Sparsity of clutter in frequency-angle domain

As in [9], denote $v_a$ as the speed of the aircraft, then the angle and the Doppler frequency of a static clutter patch satisfy the following relationship, corresponding to a ‘clutter ridge’.

$$\frac{f}{PRF} = \mu \frac{d}{\lambda} \sin \theta$$  

(8)

where

$$\mu = \frac{2v_a}{d \cdot PRF}. \quad \text{(9)}$$

The slope of the clutter ridge in frequency-angle domain is thus $\mu d \cdot PRF$.

Denote $z_k = exp(j2\pi \frac{2}{\pi} \sin \theta_k)$. Then for integer $\mu$, $\Phi$ in (1) can be rewritten as in (18). It can be seen from (18) that there are only $N + (M - 1)\mu$ distinct rows in $\Phi$, while others are repeated rows. For example, if $\mu = 1$, then the row $z_1$, $z_2$, ..., $z_N$, and the row $z_1^\mu$, $z_2^\mu$, ..., $z_N^\mu$ are identical, implying that $\Phi$ has a rank of $N + (M - 1)\mu$. Therefore, the $N_c \gg N + (M - 1)\mu$ clutter patches can be represented in terms of the $N + (M - 1)\mu$ basis vectors that span the clutter subspace. Assuming the clutter patches are mutually independent, and independent of additive noise, yields

$$R = E\{xx^H\} = E\{(\Phi \gamma + n)(\Phi \gamma + n)^H\} = \sum_{i=1}^{N_c} E(|\gamma_i|^2) \phi(\theta_i, f_i) \phi^H(\theta_i, f_i) + E\{nn^H\} = \sum_{i=1}^{N_c} E(|\gamma_i|^2) \phi(\theta_i, f_i) \phi^H(\theta_i, f_i) + \sigma^2 I$$

$$= \Phi \Gamma \Phi^H + \sigma^2 I$$  

(10)

where $\Gamma = diag(E(|\gamma_1|^2), ..., E(|\gamma_{N_c}|^2))$  

(11)

which means that $R - \sigma^2 I$ has the same rank as $\Phi$, since $\phi(\theta_i, f_i)$, $i = 1$, ..., $N_c$ are the columns of $\Phi$. Thus by Brennan’s rule, the rank of $R$ is $r \approx N + (M - 1)\mu$ and the measurements $x$ can be expressed as

$$x = \sum_{i=1}^{r} \gamma_i \phi(\hat{\theta}_i, \hat{f}_i) + n$$  

(12)

with the clutter covariance matrix $R$

$$R = \sum_{i=1}^{r} E(|\gamma_i|^2) \phi(\hat{\theta}_i, \hat{f}_i) \phi^H(\hat{\theta}_i, \hat{f}_i) + \sigma^2 I$$  

(13)

The tilde-sign $(\hat{\cdot})$ is used to denote the true basis parameters, where $\phi(\hat{\theta}_i, \hat{f}_i)$ and $\gamma_i$ are the basis vectors spanning the clutter subspace and the corresponding complex coefficients, respectively. Equation (12) indicates that each measurement can be decomposed into $r$ steering vectors of angle $\hat{\theta}_i$ and frequency $\hat{f}_i$, which shows that the clutter is sparse with a rank of $r \ll MN$ in the frequency-angle domain.
III. COMPRESSIVE SENSING WITH DICTIONARY LEARNING

A. Clutter covariance matrix estimation with compressive sensing

Compressive sensing (CS) is a well-known technique to reconstruct a sparse $N_1$-dimensional signal vector $s$ from an $N_2$-dimensional representation $x = \Phi s$ ($N_2 \ll N_1$) [12][13], where $\Phi$ is the projection matrix. In the scenario of STAP algorithm, $\Phi$ is a matrix whose columns include the basis steering vectors of the clutter subspace as in (12). CS theory proved that given $\Phi$ satisfies the Restricted Isometry Property (RIP) defined below, $s$ can be uniquely reconstructed from $x$.

**Definition 1: Restricted Isometry Property** A matrix $\Phi$ satisfies RIP with parameter $(k, \delta_k)$ if

$$(1 - \delta_k)\|s\|_2^2 \leq \|\Phi s\|_2^2 \leq (1 + \delta_k)\|s\|_2^2$$

for all $k$-sparse vectors $s$, where $k$-sparsity means there are at most $k$ non-zero elements in the vector $s$.

In [6] and [7], CS is applied to estimate the clutter covariance matrix $R$ by utilizing the sparsity in angle-frequency domain. The angle and frequency axis are discretized into an $N_s \times N_d$ ($N_s = \rho_s N, N_d = \rho_d M$) grid, where $\rho_s$ and $\rho_d$ are integers. From (1), the received signal is

$$x = \Phi_{CS}\gamma_{CS} + n$$

where $\gamma_{CS}$ is a sparse vector as discussed in the previous section (since $r \ll N_s N_d$), and $\Phi_{CS}$ consists of columns of steering vectors whose angle and frequency corresponds to the intersections in the angle-frequency domain grid. CS theory proves that the sparse vector $\gamma_{CS}$ can be solved from (15) by $l_1$ minimization,

$$\text{Minimize} \ |\gamma_{CS}|_1$$

Subject to

$$|x - \Phi_{CS}\gamma_{CS}|_2 < \epsilon$$

where $\epsilon$ bounds the resulting noise in the estimation. Equation (16) is a convex optimization problem, and can be solved using standard Semi-Definite Programming (SDP) solver in polynomial time. It is proved in [14] that $\Phi_{CS}$ in equation (15) satisfies RIP. Therefore, given $\gamma_{CS}$ is a sparse vector, it can be uniquely reconstructed. For each non-zero element $\gamma_{i,CS}$ in $\gamma_{CS}$, where $i$ is the index, the $i^{th}$ column in $\Phi_{CS}$ is the corresponding estimated basis steering vector $\phi(\theta_i,CS,f_i,CS)$. $\theta_i,CS$ and $f_i,CS$ can be solved from $\phi(\theta_i,CS,f_i,CS)$, since each steering vector is constructed based on a corresponding grid point $(\theta_i,CS,f_i,CS)$ as in (4). Then the clutter covariance matrix $R$ is calculated from (13). With this method, $R$ can be solved using only $O(r)$ samples as compared to $2MN$ in SMI. However, it requires accurate knowledge of $\theta_i$ and $f_i$ to ensure every basis steering vector of the clutter subspace corresponds to an intersection on the grid. Typically, the assumed sparsifying basis is not the same as the true sparsifying basis, resulting in basis mismatch [8], because the representative clutter patches may not located exactly at the grid points. As a result, the CS reconstruction suffers from an error that grows linearly with the element-wise mismatch between the two basis matrices. Any clutter patch that lies between two grid points leads to a column-wise difference between the true basis matrix and the assumed sparsifying basis matrix, contributing to the reconstruction error.

B. Dictionary learning

To solve the basis mismatch problem, dictionary learning algorithm is adopted to estimate the difference between the true sparsifying basis matrix and the assumed sparsifying basis matrix in (16) by minimize the squared error. Assume for the $i^{th}$ clutter patch, the distance of the true $\theta_i$ and $f_i$ from the nearest grid point is $\Delta \theta_i$ and $\Delta f_i$, and the difference between the estimated coefficients via CS and true $\hat{\gamma} = [\hat{\gamma}_1 \ldots \hat{\gamma}_{N_s,N_d}]$
is $\Delta \gamma = [\Delta \gamma_1 \ldots \Delta \gamma_{N_c,N_d}]$. If a clutter patch is located at a grid point, the coefficient corresponding to the basis vector in Eq. (12) is non-zero; otherwise the entry is zero.

Therefore,

$$x = \Phi(\theta_{i,CS} + \Delta \theta, f_{i,CS} + \Delta f_{i})(\gamma_{CS} + \Delta \gamma) + n.$$ (17)

$\gamma_{CS} + \Delta \gamma$ is a sparse vector, with only $r$ non-zero elements in it.

Denote $\alpha_i = j2\pi f_{i,CS} \frac{L}{PRF}$, $\beta_i = j2\pi \frac{\sin \theta_{i,CS}}{L} \lambda_{i,CS}$, $\Delta \alpha_i = j2\pi f_{i,CS} \frac{L}{PRF} \Delta f_{i}$, $\Delta \beta_i = j2\pi \frac{\sin \theta_{i,CS} + \Delta \theta_{i}}{L} \lambda_{i,CS}$ - $j2\pi \frac{\sin \theta_{i,CS}}{L} \lambda_{i,CS}$. Then the difference between the columns of the sparsifying matrix used in (16) and the true sparsifying matrix is shown in equation (19).

Using Taylor expansion $exp(\Delta \beta_i) = 1 + \Delta \beta_i + O(\Delta \beta_i)$, (19) can be simplified to Eq. (20). Then $\Delta \alpha_i$ and $\Delta \beta_i$ are calculated by minimizing the squared error as in (21), where $f_{i,CS}$ and $\theta_{i,CS}$ are the frequency and angle resolution of the grid used in (16), respectively. Equation (8) is used as a constraint to guarantee that all the estimated clutter patches are located on the clutter ridge. Since the objective function in (21) is a quadratic form of $\Delta \alpha_i$, $\Delta \beta_i$, and $\Delta \gamma_i$, (21) can be solved with cyclic coordinate method [15].

The overall CSDL algorithm is summarized in Table I. First, the sparsifying basis matrix and the corresponding complex basis coefficients are estimated with (16). Then the dictionary learning algorithm is applied to re-estimate frequency, angle, and basis coefficients of the basis clutter patches, and a new sparsifying basis matrix is constructed using (18) based on the estimation. After that, (16) is executed again. The process continues until the estimated frequency and angle converges. Multiple snapshots are used to estimate the expectation of the power of the clutter coefficients $E(|\gamma_i|^2)$. Then the clutter covariance matrix $R$ is estimated using Eq. (13) and the estimated frequencies, angles, and the expectation of coefficient power.

### IV. Numerical Results

The proposed CSDL algorithm is compared with RGS, PCI, MWF, and CGPAMF based on the L-band of the KASSPER dataset [10]. For the KASSPER dataset, the number of antenna elements $N = 11$, and the number of pulses $M = 32$. Therefore, for a normalized $\text{SINR} \geq -3$ dB, $L \geq 2MN = 704$ snapshots are needed for SMI method, as discussed in Section II. On the other hand, using the Brennan’s rule, $r \approx 42 \leq 704$. This is shown in Fig. 2 as a sharp rank cutoff in the eigenspectrum of a clutter covariance matrix in KASSPER dataset. The provided clutter covariance matrix in KASSPER dataset captures the real-world terrain, foliage and urban/manmade structures in a region in CA, USA. Because the true covariance matrices for all the 1000 range bins are available in the dataset, SINR can be computed directly using equation (7).

The true covariance matrix for range cell 200 (i.e., the 200th range bin) in KASSPER dataset is used to generate data for multiple snapshots, and additive white noise of clutter-to-noise ratio 40 dB is added to each snapshot. For each snapshot, since the basis coefficients are usually complex Gaussian variables, measurement $x$ is modeled as colored Gaussian random variables with covariance matrix $R$ as given in the dataset plus white Gaussian noise. Then the snapshots are processed using the five algorithms. The process is repeated for 10 iterations, where in each of them, new snapshots are generated randomly, and finally the results from the 10 iterations are averaged. This is because in each iteration, both noise and the signal are generated randomly, thus multiple iterations are needed to average out the effects of the randomness, (i.e., to obtain a smooth curve). Comparison between the five algorithms is summarized in Table II.

#### A. Computational complexity

The computational complexities, i.e., the operation counts, of the five algorithms are summarized in Table III. It shows that there exists a trade-off between accuracy and computational complexity. CSDL has the highest computational complexity, where $N_d = 352$ and $N_s = 2912$ are used for KASSPER dataset. For the numerical results in this section,
\[
\phi(f_{i,CS} + \Delta f_{i}, \theta_{i,CS} + \Delta \theta_{i}) - \phi(f_{i,CS}, \theta_{i,CS}) = \\
\begin{bmatrix}
1 - 1 \\
\exp(\beta_{i} + \Delta \beta_{i}) - \exp(\beta_{i}) \\
\vdots \\
\exp((M-1)\alpha_{i} + (N-1)\beta_{i}) - \exp((M-1)\alpha_{i} + (N-1)\beta_{i})
\end{bmatrix}
\]

(19)

\[
R = \begin{bmatrix}
\phi(f_{i,CS} + \Delta f_{i}, \theta_{i,CS} + \Delta \theta_{i}) - \phi(f_{i,CS}, \theta_{i,CS}) \\
0 \\
\exp(\beta_{i})\Delta \beta_{i} \\
\vdots \\
\exp((M-1)\alpha_{i} + (N-1)\beta_{i})(\Delta \alpha_{i} + (N-1)\beta_{i})
\end{bmatrix}
\]

(20)

Minimize \( |x - \sum_{i=1}^{N,N_{d}} (\gamma_{i,CS} + \Delta \gamma_{i})\phi(f_{i,CS} + \Delta f_{i}, \theta_{i,CS} + \Delta \theta_{i})|^2 \)

Subject to
\[
|\Delta f_{i}| \leq f_{g} \\
|\Delta \theta_{i}| \leq \theta_{g} \\
\frac{f_{i,CS} + \Delta f_{i}}{PRF} = \frac{d}{\lambda} \sin(\theta_{i,CS} + \Delta \theta_{i})
\]

(21)

### Table II

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RGS</th>
<th>PCI</th>
<th>MWF</th>
<th>CGPAMF</th>
<th>CSDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side information</td>
<td>none</td>
<td>Rank ( r )</td>
<td>Rank ( r ), steering vector ( s )</td>
<td>Estimation of the order ( P )</td>
<td>Rank ( r ), clutter ridge slope ( \frac{d}{\lambda} \sin(\theta_{i,CS} + \Delta \theta_{i}) )</td>
</tr>
<tr>
<td>Accuracy</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Computational complexity</td>
<td>low</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>

### Table III

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGS</td>
<td>( O(r^2NM) )</td>
</tr>
<tr>
<td>PCI</td>
<td>( O(N^2M^3) )</td>
</tr>
<tr>
<td>MWF with ( N_{st} ) steering vectors</td>
<td>( O(rN^2M^2N_{st}) )</td>
</tr>
<tr>
<td>CGPAMF with order ( P )</td>
<td>( O(rN^2P^{1002}) )</td>
</tr>
<tr>
<td>CSDL with a grid of ( N_{d} \times N_{s} ) intersections</td>
<td>( O(rN_{d}^3N_{s}^2N + N^2M^4) )</td>
</tr>
</tbody>
</table>

\( P = 3 \) is used for CGPAMF, and \( N_{st} = 25 \) steering vectors are evaluated.

### B. Normalized SINR

The Normalized SINR is calculated as in equation (7). It is evaluated as a function of both angle and frequency, since the steering vector \( s \) in (7) is a function of angle and frequency. In the figures, SINR is plotted as a function of either angle or frequency by averaging over the other variable. In addition, SINR is plotted in dB, i.e., \( SINR_{dB} = 10\log_{10}SINR \), and \( SINR_{dB} \leq 0 \). Note if \( \hat{R} = R \) for all the frequencies and angles, then SINR is a straight line of \( SINR = 0 \) dB. As shown in Fig. 3 and Fig. 4, CSDL has the best performance among all the benchmarks using low sample support (16 samples). The performance of PCI is worse than RGS in this scenario, because \( r \) is larger than the number of samples \( L \) and PCI uses \( r \) basis vectors, leading to inaccurate estimations of \( r - L \) basis vectors; whereas RGS uses the basis vectors which capture 99.999% of the covariance matrix energy, thus the number of basis vectors used by RGS is \( O(min(L,r)) \). The performance for the traditional SMI method is not plotted in the figure, because it is far worse than the five reduced-rank algorithms with low sample support. For \( L = 16 \), the performance for the traditional SMI method is around \( -20 \)dB.

### V. Conclusion

The CSDL algorithm is proposed to reduce the number of samples needed for clutter subspace estimation in STAP, as well as to mitigate the basis mismatch problem. By exploiting the sparsity of clutter in angle-frequency domain, CSDL performs better than other reduced-rank algorithms, at a cost of high computational complexity.

On the other hand, in practical situations, the clutter structure may not have a sharp rank cutoff, as \( \mu \) in equation (8) may not be an integer. In addition, calibration errors, wind-induced clutter motion and inaccurate measurement of velocity all give
rise to the clutter subspace leakage problem. The robustness of the CSDL algorithm against the clutter subspace leakage problem is for future investigation.

ACKNOWLEDGEMENT

The first author thanks Jason Parker from AFRL for the discussions on dictionary learning algorithms.

REFERENCES