Modified Sparse Subspace Clustering for Radar Detection in Non-stationary Clutter

Bruno Mériaux*, Arnaud Breloy†, Chengfang Ren*, Mohammed Nabil El Korso† and Philippe Forster§

* SONDRA, CentraleSupélec, 91192 Gif-sur-Yvette, France
† Paris-Nanterre University/LEME, 92410 Ville d'Avray, France
§ Paris-Nanterre University/SATIE, 94230 Cachan, France

Abstract—Detecting targets embedded in a noisy environment is an important topic in adaptive array processing. In the traditional statistical framework, this problem is addressed through a binary hypothesis test, which usually requires the estimation of side parameters from secondary data. The latter are assumed to be homogeneous and target-free, which is in practice questionable. Indeed, secondary data are usually corrupted by radar clutters and/or jammers which can be non-stationary and locally low rank. Fortunately, the latter behaviors can be well acknowledged by a union-of-subspaces model. In this work, we propose a modified subspace clustering model which can be solved using convex optimization algorithms. In the context of multiple sparse target detection and localization, a comparison is performed with various robust detection methods exhibiting advantages and drawbacks of the proposed one.

Index Terms—Target detection, Non-stationary clutter, Subspace clustering.

I. INTRODUCTION

Detecting targets enclosed in a noisy environment is a major topic in adaptive array processing, which still drives research interest. Traditionally, this problem is addressed through the statistical framework by expressing a binary hypothesis test to discriminate the presence of targets or not [1,2]. The computation of decision statistics usually requires the estimation of side parameters, such as the interference plus noise covariance matrix. This step is crucial since it directly impacts the performance of the detection process. Thus, this topic has been widely investigated through various statistical models [3–7].

However, in the classical statistical adaptive detection paradigm, the parameter learning step requires the availability of a sufficiently large dataset, called secondary data, which are assumed to be homogeneous and target-free. These conditions are not necessarily met in practice, depending on the observation environment, the measuring system, or the acquisition mode. As examples, the secondary dataset can be reduced in size, heterogeneous (i.e., non-stationary) or contaminated by targets. This issues motivated the development of numerous robust detection methods and have been addressed with various approaches, such as regularization or subspace methods [6,8–11].

Following [12], we propose in this paper to leverage robust subspace clustering techniques in order to alleviate a possible non-stationarity of the secondary dataset. This implies to reformulate the detection problem as a structured recovery/regression rather than a statistical hypothesis test (see e.g. [13]). Indeed, the whole dataset (primary and secondary data) can be modeled as the sum of sparse targets embedded locally in low rank clutter with – possibly non-stationary – interferences. In this context, recent machine learning techniques addressing the problem of subspace clustering and/or sparse matrix recovery can be used. For example, the Robust Principal Component Analysis [14] is well-suited for the recovery of "low-rank + sparse" matrices. Here, the non-stationarity of radar clutters/jammers can be better acknowledged by a union-of-subspaces model, which naturally points to the use of the subspace clustering techniques [15, 16].

In this context, [12] proposed the use of Sparse Subspace Clustering (SSC) for radar detection and illustrated its feasibility on a real dataset. Following from this work:
● We explore a variant of SSC proposed in [15] (formulating a relaxed convex problem) to include a dictionary, needed in radar detection.
● We illustrate the robustness of the SSC approach to non-stationary interference on simulated data.

This paper is organized as follows. In section II, the data model is introduced. Some limitations of the classic statistical detection are emphasized, leading to a new approach presented in Section III. In this section, we reformulate the detection problem as the recovery of a sparse matrix. In section IV, we compare the different approaches for target detection in non-stationary jamming.

In what follows, the notation $\|\mathbf{A}\|_1$ is for the ℓ_1-norm of the matrix $\mathbf{A} \in \mathbb{C}^{n \times m}$, which is given by $\|\mathbf{A}\|_1 = \sum_{n,m} |a_{n,m}|$. The operator $\text{diag}(\mathbf{A})$ collects all the diagonal elements of the matrix $\mathbf{A}$ into a vector. For a matrix $\mathbf{A}$, $\text{Tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$.

II. PROBLEM SETUP

A. Data model

Let be a radar receiver, composed by M antenna elements, collecting K snapshots, $\mathbf{z}_k \in \mathbb{C}^M$, $k = 1, \ldots, K$. Theses samples are a combination of different contributions, modeled as below [17]:

$$\mathbf{z}_k = \mathbf{v}_k + \mathbf{c}_k + \mathbf{n}_k \quad k = 1, \ldots, K \qquad \text{where,} \qquad (1)$$

● $\mathbf{v}_k$ is related to the target responses. Due to physical considerations and the geometry of the receiver, the structure of the related steering vector $\mathbf{d}$, is known. Then a dictionary $\mathbf{D}$ mapping the whole angle-of-view can be constructed as $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_T]$, with T the size of the dictionary. Thus, the vector $\mathbf{v}_k$ can be expressed as

$$\mathbf{v}_k = \mathbf{D}\boldsymbol{\alpha}_k \qquad (2)$$

where $\boldsymbol{\alpha}_k$ is the vector of power/phase shifts coefficients. In practice, only a small number of targets is active compared to the size of the dictionary. Therefore, $\boldsymbol{\alpha}_k$ is a sparse vector.
● $\mathbf{c}_k$ refers to the interferences, such as ground clutter and/or jammers. From physical consideration on the system [18], we know that the clutter contribution belongs to a low-rank subspace of size $R \ll M$. For a sufficient number of samples, i.e. $K > R$, we can then consider that a vector $\mathbf{c}_k$ can be expressed as a linear combination of the other samples, $\mathbf{c}_p$, $p \in [\![1, K]\!] \setminus \{k\}$:

$$\mathbf{C}^h \triangleq [\mathbf{c}_1, \ldots, \mathbf{c}_K] = \mathbf{C}^h \mathbf{W}^h, \text{ such that } [\mathbf{W}^h]_{i,i} = 0 \qquad (3)$$

where $\mathbf{W}^h$ sparse. The relation in (3) is also called the self-representation property of the data [15].

• $\mathbf{n}_k$ denotes a dense noise, e.g., the thermal noise, which is assumed to be white centered Gaussian distributed: $\mathbf{n}_k \sim \mathbb{CN}\left(\mathbf{0}, \sigma^2 \mathbf{I}_M\right)$.

By concatenating column-wise for all the samples, the model in (1) can be rewritten in a matrix form by

$$\mathbf{Z}^h = \mathbf{V}^h + \mathbf{C}^h + \mathbf{N}^h = \mathbf{DA}^h + \mathbf{C}^h + \mathbf{N}^h \qquad (4)$$

where the matrix $\mathbf{A}^h = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K]$ is sparse.

In the case where the whole received samples are not necessarily homogeneous, e.g., the low-rank subspace describing the interference changes during the acquisition process, the model (4) still holds. Indeed, we can consider that the heterogeneity in the samples can be modeled as an union of J unknown homogeneous subpartitions. For each homogeneous subpartition of K_j snapshots, such that $\sum_{j=1}^{J} K_j = K$, the model in (4) still remains valid. Then by concatenating column-wise each subpartition, we finally obtain

$$\mathbf{Z} = \mathbf{V} + \mathbf{C} + \mathbf{N} \text{ with } \begin{cases} \mathbf{V} = \mathbf{DA} \\ \mathbf{C} = \mathbf{CW}, \, \text{diag}\left(\mathbf{W}\right) = \mathbf{0} \end{cases} \qquad (5)$$

where $\mathbf{W}$ is a block diagonal (up to a sorting permutation) composed by the matrices $\mathbf{W}_1^h, \ldots, \mathbf{W}_J^h$. We recall that the sparse matrix $\mathbf{A}$, which is related to the present targets, is the center of interest in our detection application. However, it is worth mentioning that the sparse matrix $\mathbf{W}$, which is block-diagonal up to a sorting permutation, is useful from a clustering perspective.

B. Limitations of the classic statistical detection

Traditionally in statistical signal processing, the target detection problem can be formalized by the following binary hypothesis test:

$$\begin{cases} H_0: & \mathbf{z}_0 = \mathbf{c}_0 + \mathbf{n}_0 & ; \, \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \, \forall \, k \in [\![1, K]\!] \\ H_1: & \mathbf{z}_0 = \mathbf{D}\boldsymbol{\alpha}_0 + \mathbf{c}_0 + \mathbf{n}_0 & ; \, \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \, \forall \, k \in [\![1, K]\!] \end{cases}$$

where $\mathbf{z}_0$ is the primary sample, which is the tested sample and $\mathbf{z}_k$'s are the secondary data, which are assumed to be independant, identically distributed (i.i.d.), and target-free. Under H_0, the received signal only contains the clutter response, i.e. interference and noise. Under H_1, the tested signal additionally contains a target response. Several approaches have been proposed in the literature to design a detector [1–8, 17]. Among them, the most powerful tests are usually based on likelihood-ratio, namely the Neyman-Pearson test. In practice, the latter involve some unknown parameters of the chosen interference-plus-noise model such as the covariance matrix of the clutter. Therefore, a first step of estimation should be achieved, giving an adaptive nature to the obtained detector. To that end, the secondary data are used for the clutter's covariance estimation. The estimate plays a central role in the performance of the resulting detector, which is why this step is still driving scientific research [5, 19–21].

However, this 2-step scheme is based on two important assumptions, which can be inaccurate in practice. The first one is about the homogeneity of the secondary data, i.e. these samples are assumed to be i.i.d.. This means that either the potential non-stationarity of the clutter is not taken into account or we know a priori the partition of the secondary data in which each cluster contains i.i.d. samples. The second hypothesis concerns the target-free nature of the secondary data, which can be invalidated in the case of the presence of several targets in the observation area. From these two mentioned limitations, we propose a novel approach of the problem in order to perform a simultaneous estimation/detection process without assuming the availability of such secondary data.

III. SPARSE SUBSPACE CLUSTERING

A. Robust Subspace Recovery via bi-sparsity

In order to perform a simultaneous estimation and detection from collected observations, which are a noisy version of $\mathbf{DA} + \mathbf{C}$, we aim to recover a sparse matrix, containing the information of the targets and an union of low-rank subspaces, corresponding to non-stationary interferences. This approach, while novel for radar detection, is under a lot of ongoing investigations in machine learning or computer vision problems [15, 22]. The following minimization problem

$$\min_{\mathbf{W}, \mathbf{A}, \mathbf{C}} \|\mathbf{W}\|_1 + \lambda \|\mathbf{A}\|_1 \, \text{ s.t. } \begin{cases} \text{(i)} & \mathbf{Z} = \mathbf{DA} + \mathbf{C}, \\ \text{(ii)} & \mathbf{C} = \mathbf{CW}, \\ \text{(iii)} & \text{diag}\left(\mathbf{W}\right) = \mathbf{0} \end{cases} \qquad (6)$$

initially introduced in [16] without dictionary, and then applied with a dictionary in [12], is relevant to solve this task. Indeed, the ℓ_1-norm promotes sparsity of the matrices $\mathbf{W}$ and $\mathbf{A}$ and the parameter λ balance the two terms in the objective function. Therefore, the recovered sparse matrix $\widehat{\mathbf{A}}$ allows us to build a detection map and then to compare with a threshold the non-zero elements in order to decide the presence of the targets or not and to localize the activated atoms in the dictionary $\mathbf{D}$. Due to the bilinear constraint (ii), the problem (6) is non-convex. The optimization can be achieved by the linearized version of the Alternating Direction Method of Multipliers (ADMM) [23, 24].

B. Relaxation of the initial problem

Based on a rewriting of the problem (6) introduced in [15], we also propose to study a convexified modification of this problem. Indeed, starting from $\mathbf{Z} = \mathbf{DA} + \mathbf{C} + \mathbf{N}$ and $\mathbf{C} = \mathbf{CW}$, we obtain

$$\mathbf{ZW} = \mathbf{DAW} + \mathbf{CW} + \mathbf{NW} = \mathbf{DAW} + \mathbf{Z} - \mathbf{DA} - \mathbf{N} + \mathbf{NW}$$
$$\Rightarrow \quad \mathbf{Z} = \mathbf{ZW} + \mathbf{DA}\left(\mathbf{I} - \mathbf{W}\right) + \mathbf{N}\left(\mathbf{I} - \mathbf{W}\right)$$
$$\mathbf{Z} = \mathbf{ZW} + \mathbf{D}\tilde{\mathbf{A}} + \tilde{\mathbf{N}} \qquad (7)$$

Then, we propose to solve the following problem

$$\min_{\mathbf{W}, \tilde{\mathbf{A}}} \|\mathbf{W}\|_1 + \lambda \left\|\tilde{\mathbf{A}}\right\|_1 \, \text{ s.t. } \begin{cases} \mathbf{Z} = \mathbf{D}\tilde{\mathbf{A}} + \mathbf{ZW} \\ \text{diag}\left(\mathbf{W}\right) = \mathbf{0} \end{cases} \qquad (8)$$

where the ℓ_1-norm promotes the sparsity of the matrices $\mathbf{W}$ and $\tilde{\mathbf{A}}$ and the parameter λ balances the two terms in the criterion.

The problem (8) being convex, it can be efficiently solved using convex programming tools [25]. On one hand, the class of Accelerated Proximal Gradient algorithms, which are first-order methods, theoretically offers a convergence rate guarantee of $\mathcal{O}\left(1/k^2\right)$ [26] but needs, for this aim, an appropriate continuation technique. In practice, it can lead to variations in the convergence rate. On the other hand, the class of ADMM algorithms, which is well-suited for parallel processing, exhibits theoretically a convergence rate guarantee of $\mathcal{O}\left(1/k\right)$ [27], but offering more stability in practice. In the simulations, we chose the ADMM approach for the resolution of (8).

IV. APPLICATION TO DETECTION IN NON-STATIONARY JAMMING

In this section, we consider the problem of target detection where the interferences are due to either stationary or non-stationary jammers.
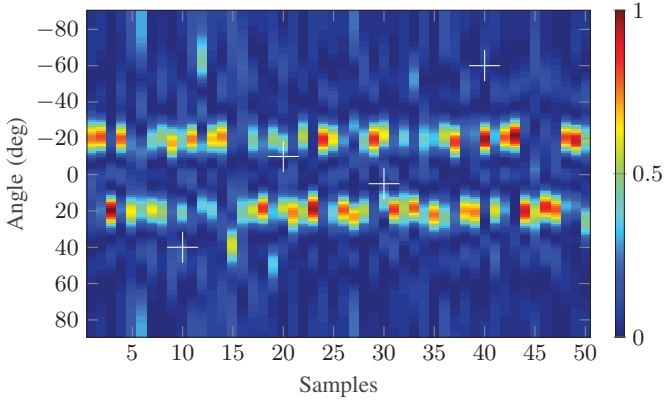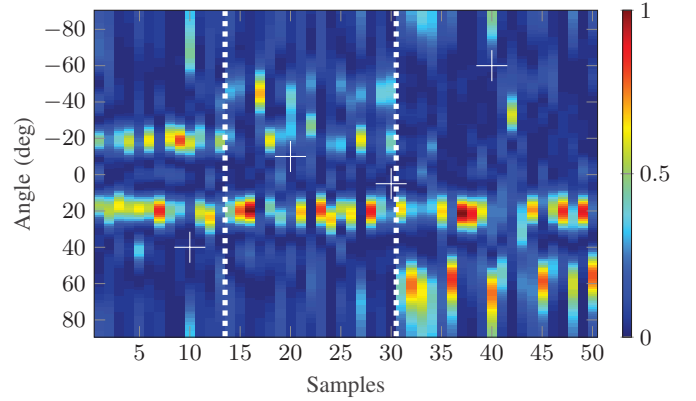
Fig. 1: Scenario with stationary jammer

Fig. 2: Scenario with non-stationary jammer

A. Case 1: stationary jammers

Let consider an uniform linear array with $M = 8$ sensors, half wavelength spaced, collecting $K = 50$ snapshots. The related steering vector is given by:

$$\mathbf{d}(\theta) = \left[1, e^{i\pi \sin \theta}, \ldots, e^{i\pi(m-1)\sin \theta}\right]^T$$

The dictionary $\mathbf{D}$ is built from $\mathbf{d}(\theta)$, for all $\theta \in [\![-90, 90[\![$. The targets of interest, which are located at $\theta_t = 40^o, 10^o, -10^o$ and -60° at different sample times, are marked as a white cross in Fig. 1. We consider in addition the presence of two jammers, through a stochastic model, in the observation scenario, which are similar as fake targets $\mathbf{d}(\theta_j)$, with $\theta_j = 20^o$ and -20°. Thus, the covariance of the jammers, which is low-rank, is given by:

$$\mathbf{R}_{\text{jam}} = \sum_{j=1}^{2} \mathbf{d}(\theta_j)\mathbf{d}(\theta_j)^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{9}$$

where $\mathbf{U}$ and $\mathbf{\Lambda}$ are the eigen-decomposition of $\mathbf{R}_{\text{jam}}$. We construct the covariance matrix of the total noise by:

$$\mathbf{R} = \frac{\text{JNR}}{\text{Tr}(\mathbf{\Gamma})}\mathbf{U}\mathbf{\Gamma}\mathbf{U}^H + \sigma^2 \mathbf{I}_M \tag{10}$$

with JNR is the Jammer to Noise Ratio and $\mathbf{\Gamma}$ is a diagonal matrix, such that $\gamma_1 = \gamma_2 = 3$. Analogously, we define the Signal to Noise Ratio (SNR) by SNR $= \frac{\|\mathbf{V}\|_2}{\sigma^2}$, where σ^2 is fixed to 1. Finally, the total noise is sampled from a centered circular complex Gaussian distribution $\mathbb{CN}(\mathbf{0}, \mathbf{R})$. In Fig. 1, a realization of the background response is plotted for all θ in the considered grid.

B. Case 2: non-stationary jammers

For the case of non-stationary jammers, we consider that in the set of $K = 50$ snapshots, there exists $J = 3$ homogeneous subpartitions, bounded by the white dotted line in Fig. 2. In the first subpartition, the jammers are in $\theta_j = 20^o$ and -20° with $\gamma_1 = \gamma_2 = 3$. In the second one, we have $\theta_j = 20^o, -45^o$ and -25° with $\gamma_1 = \gamma_3 = 2 = \gamma_2/2$. In the last one, we set $\theta_j = 60^o$ and 20° with $\gamma_1 = \gamma_3 = 3$.

C. Compared methods

In the considered application, we compare the following methods:
• the Adaptive Normalized Matched Filter (ANMF), where the covariance learning is based on the Sample Covariance Matrix, computed
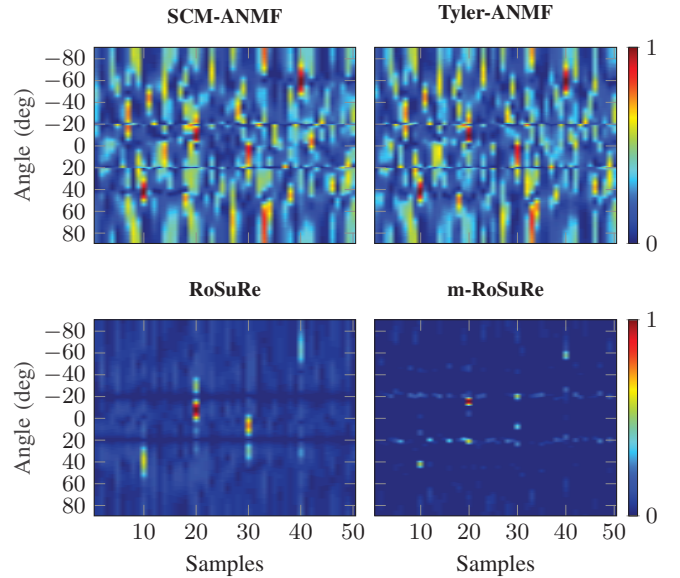

Fig. 3: Normalized detection map for case 1

from the $2M$ samples surrounding the tested sample, which will serve as secondary data.
• the ANMF, where the covariance learning is using the Tyler's estimator [28] on the $2M$ samples surrounding the tested sample.
• the RoSuRe-detector [12, 16], which solves the problem (6) by returning $\mathbf{A}$, from the given observations $\mathbf{Z}$ and the considered dictionary $\mathbf{D}$.
• the modification of RoSuRe-detector, denoted by m-RoSuRe and which is obtained by solving the problem (8) and returning $\tilde{\mathbf{A}}$, from $\mathbf{Z}$ and $\mathbf{D}$. We recall the following relation $\tilde{\mathbf{A}} = \mathbf{A}(\mathbf{I} - \mathbf{W})$.

In order to compare the different methods, the obtained detection map are insightful but not quantitative. Thus, we compute the empirical Probability of Detection (PD) for all the targets and the empirical Probability of False Alarm (PFA) (summed for all the other grid locations) with respect to a threshold. From these quantities, we can compare the Receiver Operating Characteristic (ROC) curve of each detector.
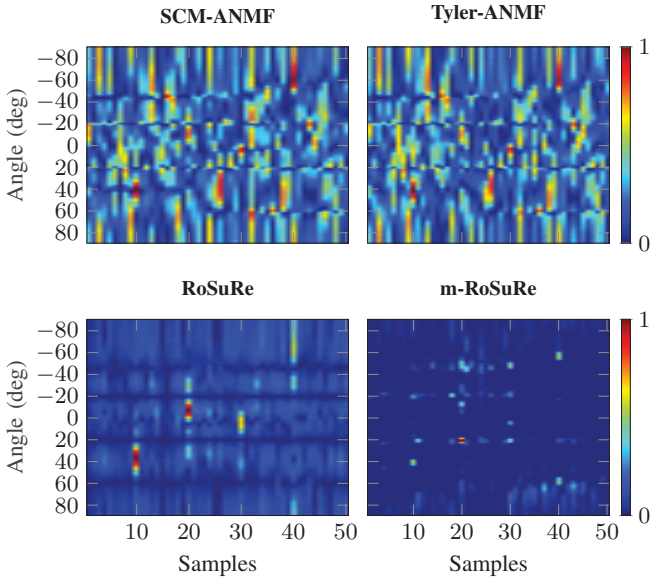
Fig. 4: Normalized detection map for case 2

Fig. 5: ROC curve: case 1 = solid line, case 2 = dash-dotted line

Fig. 6: Influence of λ: RoSuRe (solid), m-RoSuRe (dash-dot)

In Fig. 3, the normalized detection map, before any thresholding, is plotted for SNR = 9 dB and JNR = 15 dB, obtained with the different methods with a correct choice of λ for the sparse based methods. The choice of λ will be discussed later. All the methods are able to retrieve the locations and the instants of present targets. Nevertheless, the ANMF based methods introduce non negligible sidelodes, which can increase the PFA, unlike the methods using subspace clustering. In addition, we can notice that the latter remove more prominently the jammers impact. It is worth noting that the positions of the estimated and detected targets given by the modification of RoSure are slightly shifted since the target sparse matrix is different from the initial model.

In Fig. 4, we can draw the same conclusions as for Fig. 3. However, the performance of ANMF-based procedure are degraded since the assumptions on secondary data (homogeneity and target-free) are not satisfied. Furthermore, we can see that the clutter cancelation obtained by the clustering methods impacts all the samples, whereas the jammers direction changes over the time.

In Fig. 5, we can see the loss of performance for the ANMF-based methods when the jamming is non-stationary, whereas the RoSuRe-based ones seem to be less sensitive and still perform better. While having the best visual detection map, m-RoSuRe seems to reach more slowly a PD close to 1. Indeed, working with $\tilde{\mathbf{A}}$ instead of $\mathbf{A}$ may introduce some bias in the estimated target positions (recall that $\tilde{\mathbf{A}} = \mathbf{A}(\mathbf{I} - \mathbf{W})$). This phenomenon impacts the displayed detection performance as we strictly measure the PD/PFA on the actual positions.

In Fig. 6, we analyse, for the case 2, the influence of the regularization parameter λ, which balance the two ℓ_1-norms. We can see that the performance obtained by RoSuRe seems to be quite impervious to λ, unlike the one of m-RoSuRe. Again, it can be explained by the relation $\tilde{\mathbf{A}} = \mathbf{A}(\mathbf{I} - \mathbf{W})$: the dependence of the estimate on both $\mathbf{A}$ and $\mathbf{W}$ suggest that a careful selection of λ is required to mitigate the performance loss due to the introduced bias.
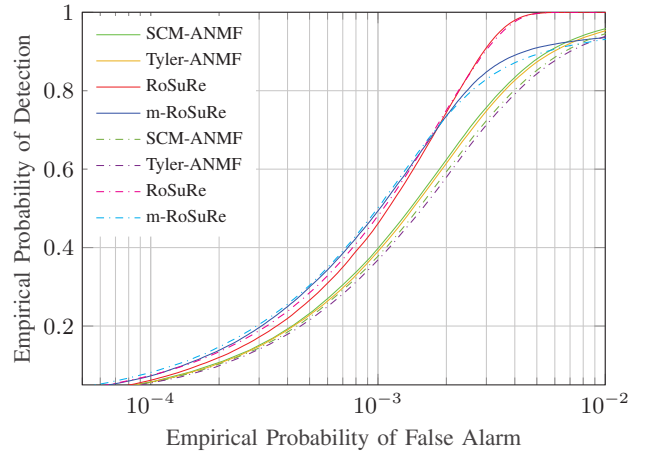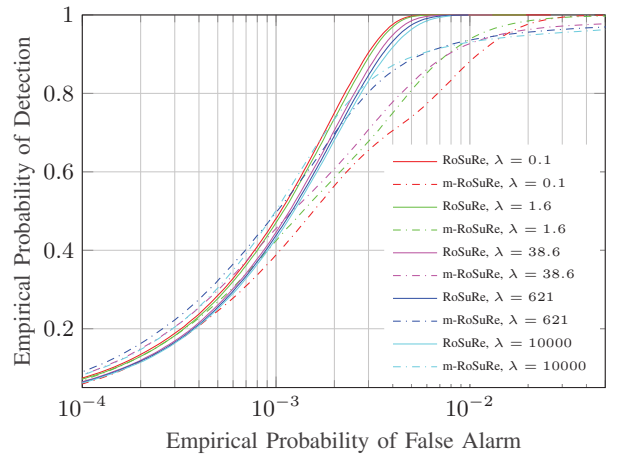
V. CONCLUSION

In this paper, we explored a new formulation for the target detection problem by means of subspace clustering approach. We proposed a convex modification of SSC that includes a dictionary. This method exhibits similar performance compared to the standard SSC but uncovers a small performance drop for a PD close to 1 due to a bias introduced on the position estimates by $\tilde{\mathbf{A}}$. A correction of this bias is left as prospect for a forthcoming study. Interestingly, theses approaches deal with all the samples at once without assumption on the interferences statistical properties. Hence, they can outperform the classic statistical detection scheme in the case of non-stationary clutter/corrupted samples.

REFERENCES

[1] E. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 2, pp. 115–127, Mar. 1986.

[2] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[3] E. Conte, M. Lops, and G. Ricci, "Adaptive detection schemes in compound-gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1058–1069, 1998.

[4] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Elsevier Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.

[5] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.

[6] A. De Maio and M. Greco, *Modern radar detection theory*. The Institution of Engineering and Technology, 2016.

[7] Y. Woodbridge, G. Elidan, and A. Wiesel, "Signal detection in complex structured para normal noise," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2306–2316, May 2017.

[8] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust statistics for signal processing*. Cambridge University Press, 2018.

[9] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[10] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 986–994, Jun. 2002.

[11] S. D. Somasundaram, N. H. Parsons, P. Li, and R. C. de Lamare, "Reduced-dimension robust capon beamforming using Krylov-subspace techniques," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 270–289, Jan. 2015.

[12] A. Breloy, M. N. El Korso, A. Panahi, and H. Krim, "Robust subspace clustering for radar detection," in *Proc. of European Signal Processing Conference (EUSIPCO)*. IEEE, 2018.

[13] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5186–5205, Aug. 2013.

[14] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, 2011.

[15] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[16] X. Bian and H. Krim, "Bi-sparsity pursuit for robust subspace recovery," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3535–3539.

[17] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Elsevier Signal Processing*, vol. 84, no. 9, pp. 1653–1665, 2004.

[18] N. A. Goodman and J. M. Stiles, "On clutter rank observed by arbitrary arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 178–186, 2006.

[19] I. Soloveychik and A. Wiesel, "Tyler's covariance matrix estimator in elliptical models with convex structure," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5251–5259, Oct. 2014.

[20] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 14, no. 64, pp. 3576–3590, Jul. 2016.

[21] A. Kammoun, R. Couillet, F. Pascal, and M.-S. Alouini, "Optimal design of the adaptive normalized matched filter detector using regularized Tyler estimators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 755–769, Apr. 2017.

[22] X. Bian, A. Panahi, and H. Krim, "Bi-sparsity pursuit: A paradigm for robust subspace recovery," *Elsevier Signal Processing*, vol. 152, pp. 148–159, 2018.

[23] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[24] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in neural information processing systems*, 2011, pp. 612–620.

[25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[26] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $\mathcal{O}\left(1/k^2\right)$," in *Soviet Math. Doklady*, vol. 27, 1983, pp. 372–376.

[27] B. He and X. Yuan, "On the $\mathcal{O}\left(1/t\right)$ convergence rate of alternating direction method," Nanjing University, Tech. Rep., 2011.

[28] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.