



**HAL**  
open science

# PERFORMANCE ANALYSIS OF A LOW-RANK DETECTOR UNDER TRAINING DATA CONTAMINATION

Pascal Vallet, Guillaume Ginolhac, Frederic Pascal, P Forster

► **To cite this version:**

Pascal Vallet, Guillaume Ginolhac, Frederic Pascal, P Forster. PERFORMANCE ANALYSIS OF A LOW-RANK DETECTOR UNDER TRAINING DATA CONTAMINATION. IEEE CAMSAP 2019, IEEE, Dec 2019, Gosier, France. 10.1109/camsap45676.2019.9022512 . hal-02388669

**HAL Id: hal-02388669**

**<https://hal.univ-grenoble-alpes.fr/hal-02388669>**

Submitted on 2 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PERFORMANCE ANALYSIS OF A LOW-RANK DETECTOR UNDER TRAINING DATA CONTAMINATION

*P. Vallet*<sup>(1)</sup>, *G. Ginolhac*<sup>(2)</sup>, *F. Pascal*<sup>(3)</sup>, *P. Forster*<sup>(4)</sup>

<sup>(1)</sup> IMS (CNRS, Univ. Bordeaux, Bordeaux INP), Talence (France)

<sup>(2)</sup> LISTIC (Univ. Savoie/Mont-Blanc, Polytech Annecy), Annecy (France)

<sup>(3)</sup> L2S (CNRS, Univ. Paris-Sud, CentraleSuplec), Gif sur Yvette (France)

<sup>(4)</sup> SATIE (CNRS, Univ. Paris-Sud, ENS Paris-Saclay), Cachan (France)

## ABSTRACT

We consider the problem of detecting a known  $M$ -dimensional target signature vector from an observation corrupted by an additive noise with unknown covariance matrix. In that case, standard statistical methods of detection usually assume that  $N$  "target free" observations are available to perform estimation of the noise covariance matrix. However, in several applications, the target signal may contaminate the training data, resulting in a deviation of the expected performance of the detectors. In this paper, we consider the performance analysis of two low-rank detectors under the assumption that  $N_c$  elements of the training data are contaminated by the target signal. More precisely, we derive the asymptotic false alarm and detection probabilities in the high dimensional regime in which both the dimension  $M$ , the number of training data  $N$  and contaminated data  $N_c$  converge to infinity at the same rate. Numerical simulations illustrate the fact that, despite the asymptotic nature of the analysis, the results obtained are accurate for reasonable values of  $M$ ,  $N$  and  $N_c$ .

**Index Terms**— LR-ANMF, Contaminated Training Data, Random Matrix Theory

## 1. INTRODUCTION

Detecting the presence of a known  $M$ -dimensional target signature vector from an observation corrupted by an additive noise is a statistical problem which arises in various fields such as radar and array processing, remote sensing, wireless communications, etc. Assuming a Gaussian model, this problem may be formulated as the following binary hypothesis test:

$$\begin{aligned}\mathcal{H}_0 : \mathbf{x} &\sim \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{R}) \\ \mathcal{H}_1 : \mathbf{x} &\sim \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \alpha^2 \mathbf{a}\mathbf{a}^* + \mathbf{R}),\end{aligned}$$

where under the null hypothesis  $\mathcal{H}_0$ , the observation vector  $\mathbf{x}$  is modeled as a complex circular Gaussian vector with covariance matrix  $\mathbf{R}$  representing the additive noise, whereas under the alternative hypothesis  $\mathcal{H}_1$ ,  $\mathbf{x}$  is modeled as a complex circular Gaussian vector with covariance matrix  $\alpha^2 \mathbf{a}\mathbf{a}^* + \mathbf{R}$ , with  $\alpha^2$  being the target response variance, and  $\mathbf{a} \in \mathbb{C}^M$  the unit-norm target signature vector. Usually, the noise covariance matrix  $\mathbf{R}$  is not directly available, and has to be estimated based on a set of  $N$  training data  $\mathbf{y}_1, \dots, \mathbf{y}_N$  by

e.g. the sample covariance matrix (SCM)

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*.$$

In that case, the observations  $\mathbf{y}_1, \dots, \mathbf{y}_N$  are assumed "target-free" and thus modeled as i.i.d.  $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{R})$  random vectors.

However, in several applications, the homogeneity assumption of the training data is unrealistic, and a certain number  $N_c$  of these data may have a covariance matrix substantially different from  $\mathbf{R}$ , in particular when the target signal contaminates the training data. Such situation may occur e.g. in radar processing where  $N$  represents the number of range/Doppler/angle cells in an area surrounding the cell under test, and where multiple target returns are located in  $N_c$  of those adjacent cells, or in array processing where  $N$  represents the number of time samples and where a potential source is transmitting during only  $N_c$  samples. In that case, the training data may be modeled<sup>1</sup> such that  $\mathbf{y}_1, \dots, \mathbf{y}_{N_c}$  are i.i.d.  $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \rho^2 \mathbf{a}\mathbf{a}^* + \mathbf{R})$  and  $\mathbf{y}_{N_c+1}, \dots, \mathbf{y}_N$  are i.i.d.  $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{R})$ , where  $\rho^2$  represents the contamination variance. We also mention that homogeneity assumption may be violated due to noise covariance mismatch [1, 2], but we do not consider this scenario here.

Unlike the standard case of non-contaminated training data, for which the theoretical performance analysis of various test statistics has been derived, the case of contaminated data seems to have received much less attention. The contamination of training data by target signal implies in general a substantial degradation of performance for standard detectors because, loosely speaking, the target signature becomes now partly assimilated to the noise covariance, resulting in a decrease of the probability of false alarm (PFA) and thus a decrease of the probability of detection (PD). The reference work [3] on the topic considers the Generalized Likelihood Ratio Test (GLRT) statistic and derives the exact PFA and PD, under the special case  $N_c = 1$ . The main difficulty in the approach comes from the fact that the GLR is an intricate statistic of  $\mathbf{x}$  and  $\mathbf{y}_1, \dots, \mathbf{y}_N$ , which requires the evaluation of various joint distributions. Other works on the topic include the derivation of improved estimators of  $\mathbf{R}$  by censoring outlier samples [4] or via Bayesian modeling [5], and non-probabilistic or experimental approaches for performance evaluation such as [6–8].

In this paper, we consider the "clutter + noise" covariance model defined as  $\mathbf{R} = \mathbf{\Gamma} + \sigma^2 \mathbf{I}$ , where  $\sigma^2$  represents a background noise

<sup>1</sup>This work has been partially supported by DGA under grant ANR-17-ASTR-0015.

<sup>1</sup>We choose here to sort the training data indexes  $\mathbf{y}_1, \dots, \mathbf{y}_N$  in the order contaminated to non-contaminated for ease of reading and without loss of generality, since  $\hat{\mathbf{R}}$  is invariant to any indexes permutation.

variance and  $\mathbf{\Gamma}$  models the covariance matrix of a clutter contribution, which is assumed rank-deficient such that

$$\mathbf{\Gamma} = \sum_{k=1}^K \gamma_k \mathbf{u}_k \mathbf{u}_k^*,$$

with  $K < M$ ,  $\gamma_1 \geq \dots \geq \gamma_K > 0$  and where  $\mathbf{u}_1, \dots, \mathbf{u}_K \in \mathbb{C}^M$  are orthonormal vectors. Under this covariance model, we focus on the performance analysis of the low-rank ANMF (Adaptive Normalized Matched Filter) test statistic derived in [9] and given by:

$$T = \frac{2M \left| \mathbf{b}^* \hat{\mathbf{\Pi}} \mathbf{x} \right|^2}{\left\| \hat{\mathbf{\Pi}} \mathbf{b} \right\|_2^2 \left\| \hat{\mathbf{\Pi}} \mathbf{x} \right\|_2^2}, \quad (1)$$

where  $\mathbf{b} \in \mathbb{C}^M$  is a unit-norm test signature vector,  $\hat{\mathbf{\Pi}}$  is the orthogonal projection matrix onto the eigenspace of  $\hat{\mathbf{R}}$  associated with the  $M - K$  smallest eigenvalues. Under the non-contaminated model, the performance of  $T$  in terms of PFA and PD has been derived in the asymptotic regime where  $N \rightarrow \infty$  while  $M$  is kept fixed. Since in several practical scenarios, the number  $N$  of observations may be constrained to be of the same order magnitude than the dimension  $M$ , an alternative asymptotic regime was proposed in [10], where both  $M$  and  $N$  are assumed to converge to infinity at the same rate, while the clutter rank  $K$  is kept fixed. In this high dimensional regime, it was shown that the LR-ANMF does not have an asymptotically constant false alarm rate (CFAR), and an alternative test statistic, which we denote  $\tilde{T}$  for the remainder, was derived and proved to retrieve the CFAR property in this double asymptotic regime.

To the best of our knowledge, there is no study on the performance evaluation of  $T$  (and a fortiori  $\tilde{T}$ ) for the training data contamination model, and we derive in this paper the asymptotic PFA and PD of both  $T$  and  $\tilde{T}$ , under the high dimensional regime mentioned above. To that purpose, we study in Section 2 the behaviour of the eigenvalues and eigenvectors of the SCM  $\hat{\mathbf{R}}$  by extending the study of [11]. Depending on the alignment of the target signature  $\mathbf{a}$  with respect to the clutter subspace  $\text{span}(\mathbf{u}_1, \dots, \mathbf{u}_K)$ , several interesting phenomena are exhibited. In Section 3, simple closed-form expressions of the asymptotic PFA and PD are provided, and numerical evaluations in Section 4 confirm the accuracy of the predicted results for realistic values of  $M$ ,  $N$  and  $N_c$ .

## 2. SPECTRAL BEHAVIOUR OF THE SCM WITH CONTAMINATED SAMPLES

From now on, we consider as in [10] the high dimensional asymptotic regime in which  $M = M(N)$  and  $N_c = N_c(N)$  are functions of  $N$  such that

$$\frac{M}{N} \xrightarrow{N \rightarrow \infty} c \in (0, 1), \quad \text{and} \quad \frac{N_c}{N} \xrightarrow{N \rightarrow \infty} \kappa \in (0, 1),$$

while  $K, \gamma_1, \dots, \gamma_K$  are considered fixed with respect to  $N$ . We further assume that  $\gamma_1 > \dots > \gamma_K > 0$  and that for all  $k = 1, \dots, K$ ,

$$|\mathbf{a}^* \mathbf{u}_k|^2 \xrightarrow{N \rightarrow \infty} \tau_k.$$

<sup>2</sup>The two assumptions  $c \in (0, 1)$  and eigenvalues of  $\mathbf{\Gamma}$  having multiplicity one are made here to keep light notations and presentation, but could be relaxed to  $c \in (0, +\infty)$  and multiplicity greater than one.

We also set  $\tau = \sum_{k=1}^K \tau_k \in [0, 1]$  and  $\mathcal{T} = \{k : \tau_k \neq 0\}$ , the latter representing the set of indexes  $k$  for which the target signature projects a non-vanishing energy onto the clutter eigendirection  $\mathbf{u}_k$ .

Let us define

$$\phi(w) = w \left( 1 - \frac{\sigma^2 c}{\sigma^2 - w} \right) \quad (2)$$

and

$$\psi(w) = \frac{(w - \sigma^2)^2 - \sigma^4 c}{(w - \sigma^2)(w - \sigma^2 + \sigma^2 c)}. \quad (3)$$

In the case of non-contaminated training data, we recall the following well-known result concerning the behaviour of the eigenvalues and eigenvectors of the SCM  $\hat{\mathbf{R}}$ , which we denote by  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_M$  and  $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M$  respectively.

**Theorem 1** ([11]). *Assume that  $\rho^2 = 0$  and that  $\gamma_1 > \dots > \gamma_K > \sigma^2 \sqrt{c}$ . Then for all  $k = 1, \dots, K$ ,*

$$\hat{\lambda}_k \xrightarrow[N \rightarrow \infty]{a.s.} \phi(\gamma_k + \sigma^2),$$

while  $\hat{\lambda}_{K+1} \rightarrow \sigma^2(1 + \sqrt{c})^2$  and  $\hat{\lambda}_M \rightarrow \sigma^2(1 - \sqrt{c})^2$  a.s. as  $N \rightarrow +\infty$ . Moreover, for all deterministic unit-norm  $\mathbf{b} \in \mathbb{C}^M$ ,

$$|\mathbf{b}^* \hat{\mathbf{u}}_k|^2 = \psi(\gamma_k + \sigma^2) |\mathbf{b}^* \mathbf{u}_k|^2 + o(1) \quad a.s.$$

Let us give some comments about the results of Theorem 1, in preparation for the statement of Theorem 2 below. For the non-contaminated model (i.e.  $\rho^2 = 0$ ),  $\mathbb{E}[\hat{\mathbf{R}}] = \mathbf{R}$  is a rank  $K$  perturbation of  $\sigma^2 \mathbf{I}$ , and we observe from Theorem 1 that each "clutter" eigenvalue  $\gamma_k, k = 1, \dots, K$  asymptotically pulls out an eigenvalue  $\hat{\lambda}_k$  of  $\hat{\mathbf{R}}$  out of the "noise" eigenvalues bulk  $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$ , assuming  $\gamma_k, k = 1, \dots, K$  are large enough (i.e. above the threshold  $\sigma^2 \sqrt{c}$ ). In this context,  $\hat{\lambda}_1, \dots, \hat{\lambda}_K$ , which are usually called *spikes*, follow in some sense the behaviour of  $\gamma_1, \dots, \gamma_K$ .

For the contaminated model, matrix

$$\mathbb{E}[\hat{\mathbf{R}}] = (1 - \kappa) \mathbf{R} + \kappa \rho^2 \mathbf{a} \mathbf{a}^*,$$

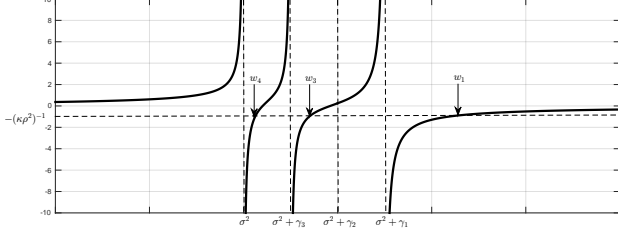
is itself a rank 1 perturbation of  $\mathbf{R}$  and intuitively, it is again expected that some eigenvalues among  $\hat{\lambda}_1, \dots, \hat{\lambda}_K$  will be shifted compared to the non-contaminated model. Moreover, the additional eigenvalue  $\hat{\lambda}_{K+1}$  may also be pulled out from the noise eigenvalues bulk. Theorem 2 below gives the precise perturbation which occurs; of course, this behaviour cannot be deduced from Theorem 1 due to the non-stationarity inherent to the contaminated model, which requires a particular treatment.

Define now the function

$$\chi(w) = \sum_{k \in \mathcal{T}} \frac{\tau_k}{\gamma_k + \sigma^2 - w} + \frac{1 - \tau}{\sigma^2 - w}, \quad (4)$$

where a typical representation is depicted in Figure 1. If  $\tau > 0$ , the equation  $\chi(w) = -(\kappa \rho^2)^{-1}$  admits  $|\mathcal{T}|$  real solutions located in the interval  $(\gamma_{\min(\mathcal{T})} + \sigma^2, +\infty)$ , and if  $\tau < 1$ , there exists an additional solution located in the interval  $(\sigma^2, \gamma_{\min(\mathcal{T})} + \sigma^2)$ . Hence, if  $\tau > 0$ , we denote by  $\omega_k$  the smallest solution greater than  $\gamma_k + \sigma^2$  for  $k \in \mathcal{T}$ , and if  $\tau < 1$ , we denote by  $\omega_{K+1}$  the smallest solution greater than  $\sigma^2$ .

Then we have the following result concerning the behaviour of the eigenvalues and eigenvectors of  $\hat{\mathbf{R}}$ , whose proof, inspired from [11], is omitted due to space constraints.



**Fig. 1.** Typical representation of  $w \mapsto \chi(w)$  for  $K = 3$ ,  $\tau_1, \tau_3 > 0$ ,  $\tau_2 = 0$  and  $\tau = \tau_1 + \tau_3 < 1$ .

**Theorem 2.** Assume that  $\rho^2 > 0$  and  $\gamma_1 > \dots > \gamma_K > \sigma^2 \sqrt{c}$ . Then for all  $k = 1, \dots, K$ ,

$$\hat{\lambda}_k \xrightarrow[N \rightarrow \infty]{a.s.} \begin{cases} \phi(w_k) & \text{if } k \in \mathcal{T} \\ \phi(\gamma_k + \sigma^2) & \text{if } k \notin \mathcal{T} \end{cases}$$

Moreover, if  $\tau < 1$ ,

$$\hat{\lambda}_{K+1} \xrightarrow[N \rightarrow \infty]{a.s.} \begin{cases} \phi(w_{K+1}) & \text{if } w_{K+1} > \sigma^2(1 + \sqrt{c}) \\ \sigma^2(1 + \sqrt{c})^2 & \text{if } w_{K+1} \leq \sigma^2(1 + \sqrt{c}) \end{cases}$$

while  $\hat{\lambda}_{K+2} \rightarrow \sigma^2(1 + \sqrt{c})^2$ , and if  $\tau = 1$ , then  $\hat{\lambda}_{K+1} \rightarrow \sigma^2(1 + \sqrt{c})^2$ , a.s. as  $N \rightarrow \infty$ . Finally, for all deterministic unit-norm vector  $\mathbf{b}, \mathbf{d} \in \mathbb{C}^M$  and  $k \in \mathcal{T}$ ,

$$\mathbf{b}^* \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* \mathbf{d} = \psi(\omega_k) \mathbf{b}^* \mathbf{v}_k \mathbf{v}_k^* \mathbf{d} + o(1) \quad a.s. \quad (5)$$

with  $\mathbf{v}_k = \frac{(\mathbf{R} - \omega_k \mathbf{I})^+ \mathbf{a}}{\|(\mathbf{R} - \omega_k \mathbf{I})^+ \mathbf{a}\|_2}$ , whereas for  $k \in \{1, \dots, K\} \setminus \mathcal{T}$ ,

$$\mathbf{b}^* \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* \mathbf{d} = \psi(\gamma_k + \sigma^2) \mathbf{b}^* \mathbf{u}_k \mathbf{u}_k^* \mathbf{d} + o(1) \quad a.s. \quad (6)$$

Recall that  $\mathcal{T}$  represents the set of indexes  $k$  for which the target vector  $\mathbf{a}$  projects a non-vanishing energy onto clutter eigenvector  $\mathbf{u}_k$ . Then Theorem 2 shows that among the spikes, only those in the set  $\{\hat{\lambda}_k : k \in \mathcal{T}\}$  will be (asymptotically) affected by the contamination, the asymptotic limit  $\phi(\gamma_k + \sigma^2)$  in the non-contaminated model being shifted upward to  $\phi(\omega_k)$  in the contaminated model.

Concerning the eigenvectors associated with the shifted spikes, the behaviour also differs significantly from the non-contaminated case. Indeed, from Theorem 1, we see that in the non-contaminated model, each eigenvector  $\hat{\mathbf{u}}_k$ ,  $k = 1, \dots, K$  projects a non-vanishing energy only onto the corresponding  $\mathbf{u}_k$  and the noise subspace  $\text{span}(\{\mathbf{u}_{K+1}, \dots, \mathbf{u}_M\})$ . In the contaminated case, the eigenvectors  $\hat{\mathbf{u}}_k$ , for  $k \in \mathcal{T}$ , will asymptotically spread onto  $\text{span}(\{\mathbf{u}_\ell : \ell \in \mathcal{T}\})$  and  $\text{span}(\{\mathbf{u}_{K+1}, \dots, \mathbf{u}_M\})$ .

**Remark 1.** Observe that for  $k \in \mathcal{T}$ ,

$$|\mathbf{b}^* \mathbf{v}_k|^2 = \frac{1}{\chi'(\omega_k)} \left| \sum_{\ell \in \mathcal{T}} \frac{\mathbf{b}^* \mathbf{u}_\ell \mathbf{u}_\ell^* \mathbf{a}}{\gamma_\ell + \sigma^2 - \omega_k} + \frac{\mathbf{b}^* \mathbf{\Pi} \mathbf{a}}{\sigma^2 - \omega_k} \right|^2 + o(1)$$

as  $N \rightarrow \infty$ . If either  $\rho^2 \rightarrow 0$  or  $\kappa \rightarrow 0$ , then for  $k \in \mathcal{T}$ ,  $\omega_k \rightarrow \gamma_k + \sigma^2$  and

$$\frac{1}{\chi'(\omega_k)} \left| \sum_{\ell \in \mathcal{T}} \frac{\mathbf{b}^* \mathbf{u}_\ell \mathbf{u}_\ell^* \mathbf{a}}{\gamma_\ell + \sigma^2 - \omega_k} + \frac{\mathbf{b}^* \mathbf{\Pi} \mathbf{a}}{\sigma^2 - \omega_k} \right|^2 \rightarrow |\mathbf{b}^* \mathbf{u}_k|^2,$$

and we retrieve, in some sense, the eigenvectors behaviour predicted in Theorem 1.

<sup>3</sup>Note that  $\{\mathbf{v}_k : k \in \mathcal{T}\}$  forms a set of unit norm, but not orthogonal, vectors.

**Remark 2.** If  $\tau < 1$  and  $\omega_{K+1} > \sigma^2(1 + \sqrt{c})$  (which occurs if  $\rho^2$  is large enough), then the sample eigenvalue  $\hat{\lambda}_{K+1}$  will split from the noise eigenvalues bulk  $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$ . Thus, any estimator of the clutter subspace dimension based on the eigenvalues  $\hat{\lambda}_1, \dots, \hat{\lambda}_M$  and consistent in the non-contaminated model may suffer from an asymptotic bias in that case.

### 3. PERFORMANCE ANALYSIS OF AN IMPROVED LR-ANMF

We study the asymptotic distribution under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  of both the LR-ANMF [9] and improved LR-ANMF [10] in the large dimensional regime, under the contaminated model for training data. The proofs of Propositions 1 and 2 below follow verbatim the one of [10, Th.1], and are therefore omitted.

For ease of reading, we define

$$\mathbf{w}_k = \begin{cases} \mathbf{v}_k & \text{if } k \in \mathcal{T} \\ \mathbf{u}_k & \text{if } k \notin \mathcal{T} \end{cases}$$

and

$$\mu_k = \begin{cases} \omega_k & \text{if } k \in \mathcal{T} \\ \gamma_k + \sigma^2 & \text{if } k \notin \mathcal{T} \end{cases},$$

so that for all  $k \in \{1, \dots, K\}$ , under the assumptions of Theorem 2,

$$\hat{\lambda}_k \xrightarrow[N \rightarrow \infty]{a.s.} \phi(\mu_k)$$

and

$$\mathbf{b}^* \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* \mathbf{d} = \psi(\mu_k) \mathbf{b}^* \mathbf{w}_k \mathbf{w}_k^* \mathbf{d} + o(1)$$

with probability one as  $N \rightarrow \infty$ . We also set

$$\Xi = \mathbf{I} - \sum_{k=1}^K \psi(\mu_k) \mathbf{w}_k \mathbf{w}_k^*.$$

**Proposition 1.** Assume that  $\rho^2 > 0$  and  $\gamma_1 > \dots > \gamma_K > \sigma^2 \sqrt{c}$ . Then,

$$\frac{T}{\vartheta} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \chi^2(2)$$

where

$$\vartheta = \frac{\sigma^2 \mathbf{b}^* \Xi \mathbf{b} + \sum_{k=1}^K \gamma_k |\mathbf{b}^* \Xi \mathbf{u}_k|^2 + \alpha^2 |\mathbf{b}^* \Xi \mathbf{a}|^2}{(\sigma^2 + \frac{\Delta}{M}) \mathbf{b}^* \Xi \mathbf{b}},$$

and with  $\Delta = \alpha^2 \mathbf{a}^* \Xi \mathbf{a} + \sum_{k=1}^K \gamma_k \mathbf{u}_k^* \Xi \mathbf{u}_k - \sigma^2 K$ .

We consider now the improved LR-ANMF derived in [10], defined as the test statistic

$$\tilde{T} = \frac{2|\mathbf{b}^* \hat{\mathbf{\Pi}} \mathbf{x}|^2}{\hat{\beta}},$$

where

$$\hat{\beta} = \hat{\sigma}^2 \left\| \hat{\mathbf{\Pi}} \mathbf{b} \right\|_2^2 + \sum_{k=1}^K (\hat{\mu}_k - \hat{\sigma}^2) \frac{(1 - \hat{\psi}(\hat{\mu}_k))^2}{\hat{\psi}(\hat{\mu}_k)} |\mathbf{b}^* \hat{\mathbf{u}}_k|^2,$$

<sup>4</sup>In fact, results are presented only for  $\mathcal{H}_1$  and may be deduced directly for  $\mathcal{H}_0$  by taking  $\alpha^2 = 0$ .

with  $\hat{\sigma}^2 = (M - K)^{-1} \sum_{k=K+1}^M \hat{\lambda}_k$ ,

$$\hat{\psi}(\mu) = \frac{(\mu - \hat{\sigma}^2)^2 - \hat{\sigma}^4 c}{(\mu - \hat{\sigma}^2)(\mu - \hat{\sigma}^2 + \hat{\sigma}^2 c)},$$

and where we have defined  $\hat{\mu}_k$  to be the unique solution to the equation  $\mu(\mu - \hat{\sigma}^2(1 - c)) = \hat{\lambda}_k(\mu - \hat{\sigma}^2)$  in the interval  $(\hat{\sigma}^2(1 + \sqrt{c}), +\infty)$  if it exists, or  $\hat{\mu}_k = \hat{\sigma}^2(1 + \sqrt{c})$  otherwise.

**Proposition 2.** Assume that  $\rho^2 > 0$  and  $\gamma_1 > \dots > \gamma_K > \sigma^2 \sqrt{c}$ . Then,

$$\frac{\tilde{T}}{\tilde{\vartheta}} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \chi^2(2)$$

where

$$\tilde{\vartheta} = \frac{\sigma^2 \mathbf{b}^* \Xi \mathbf{b} + \sum_{k=1}^K \gamma_k |\mathbf{b}^* \Xi \mathbf{u}_k|^2 + \alpha^2 |\mathbf{b}^* \Xi \mathbf{a}|^2}{\sigma^2 \mathbf{b}^* \Xi \mathbf{b} + \sum_{k=1}^K (\mu_k - \sigma^2)(1 - \psi(\mu_k))^2 |\mathbf{b}^* \mathbf{w}_k|^2}.$$

#### 4. NUMERICAL RESULTS

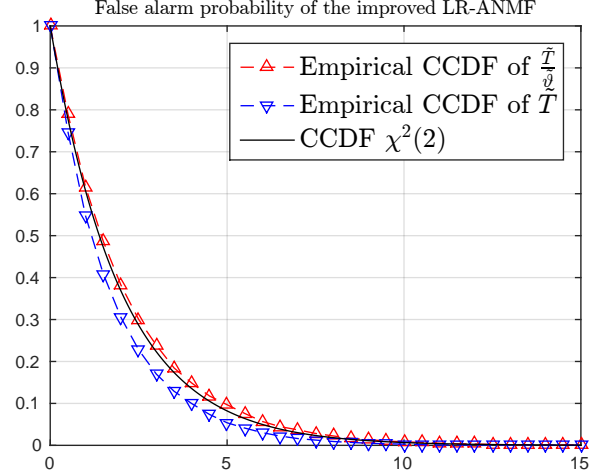
To illustrate the results of the previous sections, we consider an array processing scenario where we have a uniform linear array of  $M$  sensors with a spacing of half the wavelength, and thus the signature vector takes the form

$$\mathbf{a} = \mathbf{a}(\theta_t) = M^{-1/2} \left( 1, e^{-i\pi \sin(\theta_t)}, \dots, e^{-i(M-1)\pi \sin(\theta_t)} \right)^T,$$

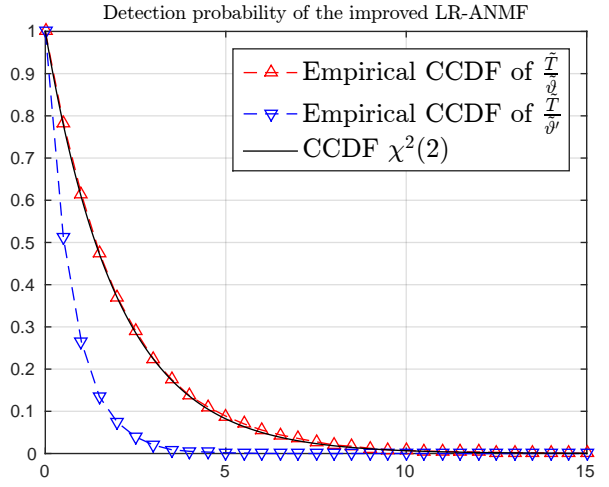
where  $\theta_t$  is the target DoA, fixed to  $20.1^\circ$ . Similarly, we have  $\mathbf{b} = \mathbf{b}(\theta)$  where  $\theta$  is the DoA under test. The clutter is composed by 3 sources with DoA  $-20^\circ, 0^\circ, 20^\circ$  and variances 0.2, 0.3, 0.5. The noise variance  $\sigma^2$  is fixed through the Clutter to Noise Ratio defined as  $\text{CNR} = \frac{1}{\sigma^2}$ , the target variance  $\alpha^2$  through the Signal to Noise Ratio  $\text{SNR} = \frac{\alpha^2}{1 + \sigma^2}$  while the contamination variance  $\rho^2$  is fixed through the Outlier to Noise Ratio as  $\text{ONR} = \frac{\rho^2}{1 + \sigma^2}$ . In the following, we only consider  $\tilde{T}$ , but the conclusions are similar for  $T$ .

In Figure 2, we consider the PFA of the improved LR-ANMF by plotting the empirical Complementary Cumulative Distribution Function (CCDF) of  $\frac{\tilde{T}}{\tilde{\vartheta}}$  under  $\mathcal{H}_0$ . We notice that the high dimensional asymptotic approximation predicted by Proposition 1 is quite accurate. In [10, Th. 1], it was proved that  $\tilde{T}$  is asymptotically  $\chi^2(2)$  distributed under the non-contaminated model, and we also plot in Figure 2 the empirical CCDF of  $\tilde{T}$ . When contamination is present, we notice that the CCDF of  $\tilde{T}$  is below the  $\chi^2(2)$  CCDF (i.e. the predicted asymptotic CCDF in the non-contaminated model). Thus contamination of the training data induces a decrease of the false alarm probability, which is intuitively expected since the signal target is now incorporated in the covariance of the training samples.

In Figure 3, we consider the PD of the improved LR-ANMF by plotting the empirical CCDF of  $\frac{\tilde{T}}{\tilde{\vartheta}}$  under  $\mathcal{H}_1$ . The empirical CCDF of  $\frac{\tilde{T}}{\tilde{\vartheta}'}$  is also plotted, where  $\tilde{\vartheta}'$  corresponds to the limit of  $\tilde{\vartheta}$  when  $\rho^2 \rightarrow 0$ , i.e.  $\tilde{\vartheta}'$  is the renormalization under the non-contaminated model such that  $\frac{\tilde{T}}{\tilde{\vartheta}'}$  is asymptotically  $\chi^2(2)$  distributed. Again, we notice as expected that the contamination of the training data implies a decrease of the detection probability, compared to what happens in the non-contaminated case.



**Fig. 2.** PFA of the improved LR-ANMF with  $M = 60$ ,  $N = 120$ ,  $N_c = 60$ ,  $\theta = 20.1^\circ$ ,  $\text{CNR} = 10$  dB,  $\text{SNR} = 5$  dB and  $\text{ONR} = 5$  dB.



**Fig. 3.** PD of the improved LR-ANMF with  $M = 60$ ,  $N = 120$ ,  $N_c = 60$ ,  $\theta = 20.1^\circ$ ,  $\text{CNR} = 10$  dB,  $\text{SNR} = 5$  dB and  $\text{ONR} = 5$  dB.

#### 5. CONCLUSION

In this paper, we have addressed the performance analysis of two low-rank detectors in a scenario where the training data used to estimate noise statistical properties are contaminated by the target signal. Both the PFA and PD have been studied in the asymptotic regime where the number of training samples and the dimension of the observations both converge to infinity at the same rate. Numerical results confirm the accuracy of the asymptotic predictions for realistic values of sample size and observations dimension, and make it possible to quantify the decrease of the PFA and PD induced by the contamination.

## 6. REFERENCES

- [1] S. Bidon, O. Besson, and J.-Y. Tournet, "A Bayesian approach to adaptive detection in nonhomogeneous environments," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 205–217, 2008.
- [2] R. Raghavan, "False Alarm Analysis of the AMF Algorithm for Mismatched Training," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 83–96, 2019.
- [3] K. Gerlach, "The effects of signal contamination on two adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 297–309, 1995.
- [4] K. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 885–901, 2002.
- [5] O. Besson and S. Bidon, "Adaptive processing with signal contaminated training samples," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4318–4329, 2013.
- [6] M. Farrell and R. Mersereau, "On the impact of covariance contamination for adaptive detection in hyperspectral imaging," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 649–652, 2005.
- [7] J. Theiler and B. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 1, pp. 98–102, 2006.
- [8] S. Matteoli, M. Diani, and G. Corsini, "Impact of signal contamination on the adaptive detection performance of local hyperspectral anomalies," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1948–1968, 2013.
- [9] M. Rangaswamy, F.C. Lin, and K.R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Process.*, vol. 84, no. 9, pp. 1653–1665, 2004.
- [10] P. Vallet, G. Ginolhac, F. Pascal, and P. Forster, "An Improved Low Rank Detector in the High Dimensional Regime," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5336–5340.
- [11] F. Benaych-Georges and R.R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *J. Multivariate Anal.*, vol. 111, no. 0, pp. 120–135, 2012.