A Two-Stage Detector for Operation in Outlier-dense Scenario

Sudan Han[1], Xiaotao Huang[1], Antonio De Maio[2], Luca Pallotta[2], and Salvatore Iommelli[3]

[1] National University of Defense Technology, Changsha, China. (e-mail: xiaoxiaosu0626@163.com, xthuang@nudt.edu.cn)
[2] University of Federico II, Naples I-80125, Italy (e-mail: ademaio@unina.it, luca.pallotta@unina.it)
[3] Ente di Formazione Professionale Maxwell, Naples I-80145, Italy (e-mail: s.iommelli@entemaxwell.it)

1. Motivations

Reliable target detection based on a set of secondary data is a crucial issue in radar processing.

The presence of **heterogenous data**, especially data containing **outliers** which share similar steering vectors as the desired target, might cause a significant **performance degradation** of typical detectors.

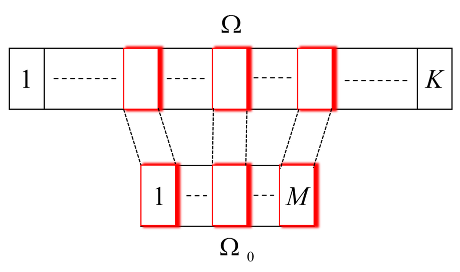
To overcome the drawback of the outliers, we devise a **two-stage detector** which first selects the most homogeneous training data from the available secondary dataset and then exploits a CFAR detector. As to the training data selection procedures, we obtain the **maximum likelihood (ML) estimate** of the outlier subset resorting to the generalized likelihood function (GLF) and remove the data vectors whose indices belong to the estimated subset. We also design an **approximate procedure** in order to reduce the computational load. Then we combine this selector together with the **adaptive matched filter (AMF)** to construct the two-stage detector.

2. Data Model

Assume a radar system collecting data from N channels (spatial and/or temporal). The returns from the range cell under test and K secondary range cells are properly sampled to form the N-dimensional **primaray data** $\mathbf{x}$ and **secondary data** $\mathbf{x}_i, i=1,\cdots,K$, respectively, and suppose that

$$\begin{cases} \mathbf{x}_i = \mathbf{c}_i, \ \forall i \in \Omega - \Omega_0 \\ \mathbf{x}_i = \mathbf{c}_i + \mathbf{p}_i, \ \forall i \in \Omega_0 \end{cases}$$

- $\mathbf{c}_1, \cdots, \mathbf{c}_K$ are independent, circular, zero-mean, complex **Gaussian** random vectors with **the same covariance matrix** $\mathbf{R}$;
- $\Omega = \{1, \cdots, K\}$ is a set of size K;
- $\Omega_0 = \{i_1, \cdots, i_M\}$ is a **subset** of Ω with distinct elements and of size M, denoting the **outlier subset**;
- $\mathbf{p}_i$s are unknown, possibly random, **outlier vectors**.



3. ML Estimate of the Outlier Subset

Resorting to the GLF and modeling $\mathbf{R}, \mathbf{p}_{i_1}, \cdots, \mathbf{p}_{i_M}$ as unknown quantities, the **ML estimate** of Ω_0 is the solution to the following optimization problem

$$\hat{\Omega}_0 = \arg\max_{\Omega_0}[\max_{\mathbf{R}} \max_{\mathbf{p}_{i_1},\cdots,\mathbf{p}_{i_M}} f(\mathbf{x}_1, \cdots, \mathbf{x}_K | \mathbf{R}, \mathbf{p}_{i_1}, \cdots, \mathbf{p}_{i_M})]$$

where $\arg\max_{\Omega_0}(\cdot)$ denotes the set between the $\binom{K}{M} = \dfrac{K!}{(K-M)!M!}$ subsets of Ω with distinct elements and of size M which maximizes the argument and $f(\mathbf{x}_1, \cdots, \mathbf{x}_K | \mathbf{R}, \mathbf{p}_{i_1}, \cdots, \mathbf{p}_{i_M})$ is the **joint probability density function (pdf)** of $\mathbf{x}_1, \cdots, \mathbf{x}_K$.

After some algerbra, the above optimization problem is tantamount to solving:

$$\hat{\Omega}_0 = \arg\min_{\Omega_0}[\det(\mathbf{R}_x)]$$

where $\mathbf{R}_x = \dfrac{1}{K-M} \sum_{i \in \Omega - \Omega_0} \mathbf{x}_i \mathbf{x}_i^\dagger$ is the sample covariance matrix (SCM) corresponding to $\Omega - \Omega_0$.

4. Approximate ML procedure

The above problem is a **combinatorial optimization** problem whose **computational burden is heavy**. It is thus of interest developing **approximate procedures** which permit **a more affordable computational load** and ensure **good quality solutions**. Toward this goal, we give the following theorem:

Theorem 1. Consider a dataset $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_K\}$ containing K random vectors of size N. Let $H_1 \subset \{1, \cdots, K\}$ with $|H_1| = h$ $(N \le h = (K-M) \le K)$ and evaluate the SCM $\mathbf{S}_1 = (1/h)\sum_{i \in H_1} \mathbf{x}_i \mathbf{x}_i^\dagger$. Then compute the GIP values for all the data

$$\mathbf{d}_1(i) = \mathbf{x}_i^\dagger \mathbf{S}_1^{-1} \mathbf{x}_i, \quad \text{for} \quad i = 1, \cdots, K$$

Now take the indices associated with the lowest h GIP values to construct H_2, and compute the new SCM $\mathbf{S}_2 = (1/h)\sum_{i \in H_2} \mathbf{x}_i \mathbf{x}_i^\dagger$ based on H_2, then $\det(\mathbf{S}_2) \le \det(\mathbf{S}_1)$ with equality if and only if $\mathbf{S}_2 = \mathbf{S}_1$.

- Starting from an initial SCM $\mathbf{S}_1$, we could obtain a **more concentrated** SCM $\mathbf{S}_2$ (i.e. sharing a lower determinant),so this procedure is referred to as **Concentration-step (C-step)**;
- An iterative algorithm which provides a sequence of secondary datasets (with cardinality h) characterized by a **non-increasing SCM determinant** can be obtained;
- The iterative algorithm must **converge** due to the finitely many h-subsets, and the **stopping criterion** can be set as $\det(\mathbf{S}_m) = \det(\mathbf{S}_{m-1})$.
- There is no guarantee that the iterative algorithm converges to **the global optimum** of the minimal covariance determinant problem.

Fortunately, if we take **more initial subsets**, apply C-steps on each subset until convergence and select the one leading to **the lowest covariance determinant**, the qualtiy of the solution improves.

To implement the above iterative procedure, it is necessary to specify how it is initialized. In this respect, we consider the following two methods:

1. Construct a random h-subset H_1 (the AML based on this initialization is referred to as **AML-h**).

2. Construct a random N-subset H and evaluate the SCM $\mathbf{S}_0 = (1/N)\sum_{i \in H} \mathbf{x}_i \mathbf{x}_i^\dagger$. Then compute the GIP values for all the data $\mathbf{d}_0(i) = \mathbf{x}_i^\dagger \mathbf{S}_0^{-1} \mathbf{x}_i, \ i = 1, \cdots K$. Sort them in ascending order and select the indices corresponding to the lowest h GIP values to form the initial h-subset H_1 (the AML procedure employing this initialization is referred to as **AML-N**).

Both the initializations share the same pseudocode which is summarized in Algorithm 1:

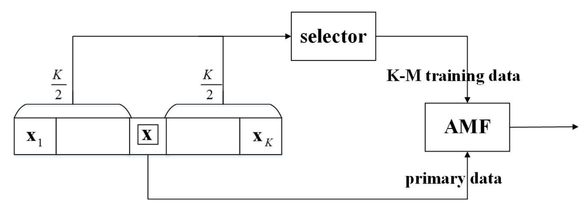
Algorithm 1: Pseudocode of the AML method

1. Construct an initial h-subset and carry out C-steps until convergence;

2. Repeat step 1 until a maximum pre-set number of initial subsets $N_{initial}$ is reached;

3. Report the subset with the lowest covariance determinant chosen among the $N_{initial}$ convergent subsets as the outlier-free dataset and the corresponding complementary set as the outlier set.

5. Analysis of the Two-stage Receiver



We consider the a two-stage receiver composed of **the training data selector based on the approximate ML procedure** plus **the AMF detector**, whose decision rule can be expressed as

$$\frac{|\mathbf{p}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}|^2}{\mathbf{p}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{p}} \underset{H_0}{\overset{H_1}{\gtrless}} T \quad \text{with} \quad \mathbf{p} = [1, e^{j2\pi f_{d_t}}, \cdots, e^{j2\pi(N-1)f_{d_t}}] \quad \text{and} \quad \hat{\mathbf{R}} = \frac{1}{K-M}\sum_{i \in \Omega - \hat{\Omega}_0} \mathbf{x}_i \mathbf{x}_i^\dagger$$

where $\mathbf{p}$ is the **steering vector of the desired target** with normalized Doppler frequency f_{d_t}, $\hat{\mathbf{R}}$ is the estimated interference covariance matrix, $\hat{\Omega}_0$ is the **estimated outlier subset** and T is the **detection threshold** which is set resorting to $10^3/P_{\text{fa}}$ (P_{fa} is **the nominal probability of false alarm**) Monte Carlo simulations assuming **homogeneous** $(K-M)$ samples at the input of the AMF detector.

We consider the following interference covariance matrix model

$$\mathbf{R} = \mathbf{R}_0 + \mathbf{I} \quad \text{with} \quad \mathbf{R}_0(i,j) = \sigma_c^2 \rho^{|i-j|} e^{j2\pi f_{d_c}(i-j)}, \quad i,j = 1, \cdots, N$$

where $\mathbf{R}_0$ accounts for the **exponentially shaped clutter**, σ_c^2 is the clutter to noise power ratio (CNR), ρ is the one-lag correlation coefficient, f_{d_c} is the normalized clutter Doppler frequency and $\mathbf{I}$ accounts for **the thermal noise**.

We randomly inject N_o outliers with **equal powers** into **different secondary range cells**. The steering vectors of these outliers are given by

$$\mathbf{p}_i = \alpha_i [1, e^{j2\pi f_{d_{o,i}}}, \cdots, e^{j2\pi(N-1)f_{d_{o,i}}}], \quad i = 1, \cdots, N_o$$

where α_i and $f_{d_{o,i}}$ are the complex amplitude and normalized Doppler frequency of the ith outlier, respectively. Moreover, $P_o = |\alpha_i|^2, i = 1, \cdots, N_o$ denotes the outlier power.

In the numerical experiments, we study **the probability of detection** P_{d} versus **the signal-to-interference-plus-noise ratio** (SINR: SINR $= |\alpha_t|^2 \mathbf{p}^\dagger \mathbf{R}^{-1}\mathbf{p}$ with α_t denoting the complex amplitude of the desired target) and **actual false alarm probability** (normalized by the nominal value) considering

$$K = 20, \ N = 8, \ N_{initial} = 40, \ P_{\text{fa}} = 10^{-4}, \ \sigma_c^2 = 20 \text{ dB}, \ \rho = 0.95, \ f_{d_c} = 0.05, \ N_o = 3, \ M = 3$$

for the cases:

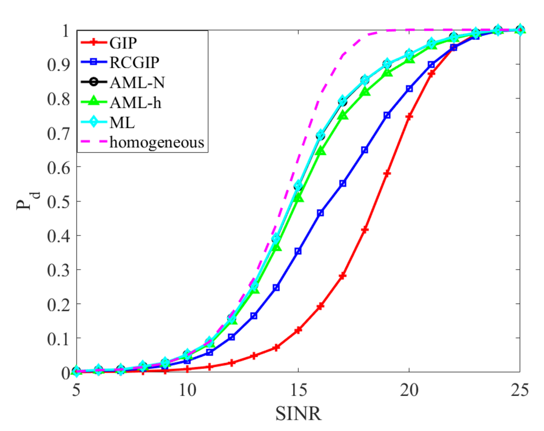
1. $f_{d_{o,i}} = 0.15, i = 1, \cdots, 3, \ P_o = 20$ dB;
2. $f_{d_{o,i}} = 0.15, i = 1, \cdots, 3, \ P_o = 30$ dB;
3. the normalized Doppler frequencies of the outliers are modeled as statistically independent random variables uniformly distributed within the interval [0.1,0.2], $P_o = 20$ dB;
4. outliers with random Doppler frequencies, $P_o = 30$ dB;

6. Conclusions

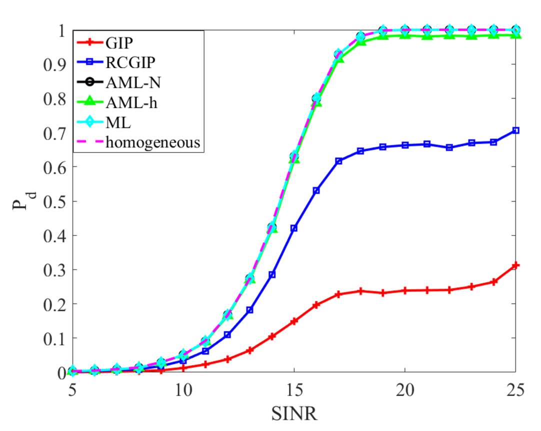
In this work, we have designed a two-stage detector to counter the presence of outliers.

- We have derived **the ML estimate** of the outlier subset;
- We have devised an **approximate ML procedure** to reduce the computational complexity;
- We have combined **the training data selector** based on the approximate ML procedure and a **CFAR detector AMF** to perform the target detection;
- We have evaluated the performance of the proposed two-stage detector based on simulated data.

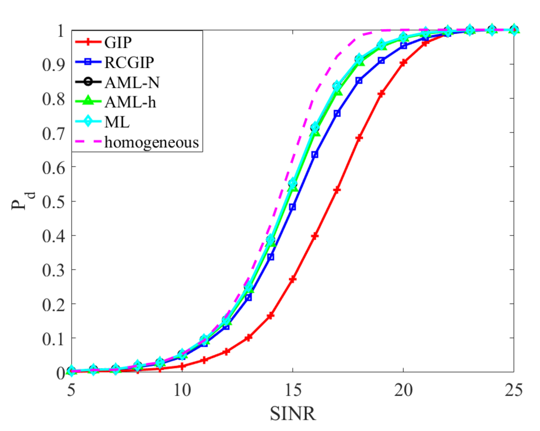
Numerical Results



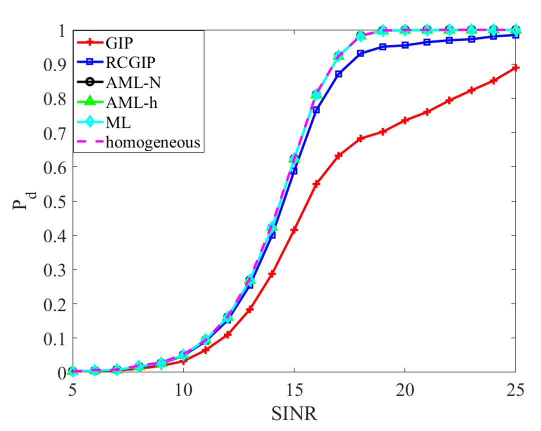
Case 1



Case 2



Case 3



Case 4

Table 1: Actual false alarm probability (normalized by the nominal value) for different receivers

Method	GIP	RCGIP	AML-N	AML-h	ML	homogeneous
case 1	0.58	1.34	1.43	1.41	1.44	1.00
case 2	0.41	0.99	1.01	1.01	1.01	1.01
case 3	0.83	1.49	1.53	1.48	1.53	1.01
case 4	0.85	1.01	0.99	0.99	0.99	0.99