Signals & Systems - Detection class

About [2] *"The CFAR Adaptive Subspace Detector is a Scale-Invariant GLRT"* (S. Kraut and L. L. Scharf. 1998)

GOMEL Jules

December 8, 2023

1 Introduction

1.1 Contextualization and Problem Statement

Detection problems revolve around determining the presence or absence of a signal-of-interest (SoI) among signals measured by sensors. The detection theory is highly versatile and finds applications across various fields, ranging from biology—for characterizing human auditory discrimination, for instance—to telecommunications, particularly in radar detection scenarii. In a detection case, four distinct situations arise based on the presence of a signal of interest among the captured signals and the detector's response.

The four results of a detection situation are summarized in Table (1), depending on the presence or not of a SoI and on the response of the detector.

	Detector Response	
	"No SoI"	**"SoI present"**
SoI present	Miss	Hit
No SoI	Correct Rejection	False Alarm

Table 1: Detection Situations

One of the most crucial steps in the history of detection theory was the introduction of threshold detectors. This concept relies on deriving a test statistic from the measured signals, comparing it to a threshold, and making a decision regarding the presence of SoI based on the outcome of the test.

Traditional detectors, such as matched filter detectors, aim to identify an *a priori* template in the measured signal when we possess knowledge about the properties of the expected SoI [6]. This fundamental detector is particularly useful for detecting a deterministic SoI in the presence of additive Gaussian noise.

However, in numerous real-world scenarios, surveillance and communication systems must operate in environments characterized by dynamic conditions. These conditions can lead to variations in various system properties, including unwanted echoes - clutter -, poor signal-to-noise ratio, or an important presence of interferences. This situation motivated the development of adaptive detectors to address scenarios where many components of the noise are unknown, situations in which traditional detectors perform pretty badly.

1.2 Brief State of the Art at the Time of the Article

The founding article of adaptive detection is the one published "An Adaptive detection algorithm" written by E.J. Kelly [1]. This article was a landmark in the history of detection theory as it was one of the very first to address a general problem of signal detection in a background of unknown Gaussian noise and thus, this article built one of the first method being adaptive as it could handle a situation of varying conditions as described before. Especially, this article tackled a situation in which the noise covariance matrix or noise structure was unknown.

However, the article "The CFAR Adaptive Subspace Detector is a Scale-Invariant GLRT" from S. Kraut and L.L. Scharf [2] at stake in this report aims to be even more general by being able to perform in a situation in which the noise scaling factor is unknown.

In order to have a better understanding of the scope of the article we will study in this report, it is important to also cite the preliminary article from L.L. Scharf from 1996 [4] which developed a detector called "Matched-Subspace Detector" or MSD which relies on using the projection of the SoI subspace on the SoI plus interferences subspace as a test statistic to be compared to a threshold, and the article we will develop today aims to show this detector is equivalent to the Generalized Likelihood Ratio Test which is broadly used in detection theory.

2 Test Statistic Calculation

2.1 Summary of the hypothesis tested

Let's describe what is the detector involved in this detector.

Given a N dimensions measurement vector $\underline{y}$ and a set $\boldsymbol{X}$ of M training vector $\underline{x}_i$, the binary hypothesis testing at stake here is the following :

$$\begin{cases} H_0 : y \sim \mathcal{CN}(\underline{0}, \sigma^2 \boldsymbol{R}) \; ; \forall i \in [\![1, M]\!], \underline{x}_i \sim \mathcal{CN}(\underline{0}, \boldsymbol{R}) \\ H_1 : y \sim \mathcal{CN}(\mu e^{j\alpha} \underline{\Psi}, \sigma^2 \boldsymbol{R}) \; ; \forall i \in [\![1, M]\!], \underline{x}_i \sim \mathcal{CN}(\underline{0}, \boldsymbol{R}) \end{cases} \tag{1}$$

In this hypothesis test, $\underline{\Psi}$ is the deterministic, complex steering vector and is known. The unknown parameter are the noise scaling $\sigma^2 \in \mathbb{R}$, the noise structure $\boldsymbol{R} \in \mathbb{R}^N$ and the signal scaling and phase $\mu e^{j\alpha} \in \mathbb{C}$. The detector described in Equation(1) is called the Adaptive Subspace Detector (ASD).

2.2 Test Statistic for the Detector

The proposed test statistic for the Adaptive Subspace Detector is the cosine-squared of the angle made by the whitened measurement vector $\boldsymbol{R}^{-1/2}\underline{y}$ in the subspace spanned by the whitened steering vector $\langle \boldsymbol{R}^{-1/2}\underline{\psi} \rangle$.

Theoretically, is defined as the ratio between the scalar product between $\boldsymbol{R}^{-1/2}\underline{y}$ and $\boldsymbol{R}^{-1/2}\underline{\psi}$ divided by the norm of these two vectors, as following :

$$\cos^2 = \frac{(\boldsymbol{R}^{-1/2}\underline{y} | \boldsymbol{R}^{-1/2}\underline{\Psi})}{(\boldsymbol{R}^{-1/2}\underline{\Psi} | \boldsymbol{R}^{-1/2}\underline{\Psi})(\boldsymbol{R}^{-1/2}\underline{y} | \boldsymbol{R}^{-1/2}\underline{y})}$$

$$\cos^2 = \frac{|\underline{\Psi}^H \boldsymbol{R}^{-1} \underline{y}|^2}{(\underline{\Psi}^H \boldsymbol{R}^{-1} \underline{\Psi})(\underline{y}^H \boldsymbol{R}^{-1} \underline{y})} \tag{2}$$

with the canonical scalar product between two vectors $\underline{u}, \underline{v}$ in $\mathbb{C}^N$ being $(\underline{u}|\underline{v}) = \underline{u}^h \underline{v}$.

However, as defined before, the noise structure $\boldsymbol{R}$ is unknown but we can exploit the set of training vectors to estimate it by computing $S = \boldsymbol{X}\boldsymbol{X}^H$ and thus, we have an estimate of the cosine-squared as follows :

$$\widehat{\cos^2} = \frac{|\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{y}|^2}{(\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{\Psi})(\underline{y}^H \boldsymbol{S}^{-1} \underline{y})} \tag{3}$$

The test statistic investigated by the article [2] is the estimate of the cosine-squared as described in Equation (3) and will be refered to as the ASD test statistic.

2.3 The GLRT is the ASD test

Let's now detail how this test statistic can be involved in a binary hypothesis testing and linked to the Generalized Likelihood Ratio Test (GLRT) of the detector described in Equation (1).

The general formula of the Generalized Likelihood Ratio (GLR) $\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y})$ for a detector is known as the ratio between the probability density function (PDF) of the detector under the alternative hypothesis H_1 noted f_1 divided by the PDF of the detector under H_0 noted f_0, with the set of unknown parameters $\left\{\boldsymbol{R}, \mu e^{j\alpha}, \sigma^2\right\}$ being estimated by their maximum likelihood estimates (MLE) noted $\left\{\hat{\boldsymbol{R}}, \widehat{\mu e^{j\alpha}}, \hat{\sigma^2}\right\}$.

The GLRT is then the comparison of this ratio to a certain threshold η ensuring a fixed false-alarm rate P_{fa}. This test can be summarized as follows in Equation(4) :

$$\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y}) = \frac{f_1(\boldsymbol{X}, \underline{y} | \hat{\boldsymbol{R}}_{0,1}, \hat{\sigma}^2_{0,1}, \widehat{\mu e^{j\alpha}})}{f_0(\boldsymbol{X}, \underline{y} | \hat{\boldsymbol{R}}_{0,1}, \hat{\sigma}^2_{0,1})} \underset{<}{\overset{>}{}}{}^{'H_1'}_{'H_0'} \eta \tag{4}$$

It is now appropriate to calculate $f_1(\boldsymbol{X}, \underline{y})$ and $f_0(\boldsymbol{X}, \underline{y})$ then compute the estimate of the parameters $\hat{\theta}$ before developing the expression of the GLR.

2.3.1 Probability density functions

The probability density function of the training dataset $\boldsymbol{X}$ and the measurement $\underline{y}$ under the alternate hypothesis H_1 can be developed as follows :

$f_1(\boldsymbol{X}, \underline{y}) = f_1(\underline{y}) \prod_{i=1}^{M} f(\underline{x}_i)$ as the training vectors $\underline{x}_i$ and $\underline{y}$ are all mutually independent.

$$f_1(\boldsymbol{X}, \underline{y}) = \frac{1}{\pi^N ||\sigma^2 \boldsymbol{R}||} \exp\left(\frac{-1}{\sigma^2}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H \boldsymbol{R}^{-1}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})\right) \prod_{i=1}^{M} \frac{1}{\pi^N ||\boldsymbol{R}||} \exp\left(\frac{-1}{\sigma^2} \underline{x}_i^H \boldsymbol{R}^{-1} \underline{x}_i\right)$$

as $\underline{x}_i \sim \mathcal{CN}(\underline{0}, \boldsymbol{R})$ and $\underline{y} \sim \mathcal{CN}(\mu e^{j\alpha}\underline{\Psi}, \sigma^2 \boldsymbol{R})$ under H_1.

We can then compute f_0 as $f_1|_{\mu=0}$. The article then rewrite this expression to a more compact form, exploiting properties of the determinant, some matrices simplifications and introducing a matrix $\boldsymbol{T}$:

$$f_{0,1} = \left\{ \frac{1}{\pi^N ||\boldsymbol{R}|| \sigma^{\frac{2N}{M+1}}} \exp\left(-\mathrm{tr}(\boldsymbol{R}^{-1}\boldsymbol{T}_{0,1})\right) \right\}^{M+1} \tag{5}$$

where

$$\begin{cases} \boldsymbol{T}_1 = \frac{1}{M+1} \left\{ (\underline{y} - \mu e^{j\alpha}\underline{\Psi})(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H + \sum_{i=1}^{M} \underline{x}_i \underline{x}_i^H \right\} \\ \boldsymbol{T}_0 = \boldsymbol{T}_1|_{\mu=0} \end{cases}$$

With the expression in Equation(5), we have a compact form of the expression of the probability densities under each hypothesis, involved in the computation of the GLR.

2.3.2 Maximum likelihood estimates of the parameters

We remind that the unknown parameters are the noise structure $\boldsymbol{R}$, the noise scaling factor σ^2, and signal scaling and phase $\mu e^{j\alpha}$. Now is the moment to compute the MLE of the parameters under each hypothesis before inserting their expressions into the Generalized Likelihood Ratio (GLR).

In order to compute the MLE of a parameter θ, one has to maximize the likelihood function according to this parameter, i.e., finding $\hat{\theta}$ such that:

$$f_{0,1}(\boldsymbol{X}, \underline{y} | \hat{\theta}) = \max_{\theta} f_{0,1}(\boldsymbol{X}, \underline{y}) \tag{6}$$

Following the computation in part 5.1, one shows that the MLE of the noise structure R under the null hypothesis H_0, noted $\hat{\boldsymbol{R}}_0$, and under the alternate hypothesis H_1, noted $\hat{\boldsymbol{R}}_1$, are the following:

$$\hat{\boldsymbol{R}}_0 = \boldsymbol{T}_0 \quad ; \quad \hat{\boldsymbol{R}}_1 = \boldsymbol{T}_1 \tag{7}$$

One then has to inject the expression of $\hat{\boldsymbol{R}}_0 = \boldsymbol{T}_{0,1}$ into $f_{0,1}$, giving the following form:

$$f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}) = \left\{ \frac{1}{(e\pi)^N ||\boldsymbol{T}_{0,1}|| \sigma^{2\frac{N}{M+1}}} \right\}^{M+1} \tag{8}$$

To minimize this expression with respect to σ^2, we used the simplification of the determinant of $\boldsymbol{T}_{0,1}$ detailed in [1], stated as follows:

$$\begin{cases} ||\boldsymbol{T}_1|| = \left(\frac{M}{M+1}\right)^N ||\boldsymbol{S}|| \left(1 + \frac{1}{M\sigma^2}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \mu e^{j\alpha})\right) \\ ||\boldsymbol{T}_0|| = ||\boldsymbol{T}_1|| \,|_{\mu=0} \end{cases} \tag{9}$$

The MLE of the noise scaling under each hypothesis is eventually expressed the following way:

$$\begin{cases} \hat{\sigma}_1^2 = \frac{M-N+1}{MN}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \mu e^{j\alpha}) \\ \hat{\sigma}_0^2 = \hat{\sigma^2}_1|_{\mu=0} \end{cases} \tag{10}$$

The computations to get the expression of $\hat{\sigma}_{0,1}^2$ are detailed in part 5.2. The next step is to put the expressions of these estimates in Equation(8) and develop it with the expressions of the determinants in Equation(9).

$$f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}, \hat{\sigma}_{0,1}^2) = \left\{ \frac{(M+1)^N}{(e\pi M)^N ||\boldsymbol{S}|| (1 + \frac{N}{M-N+1})(\hat{\sigma}_{0,1}^2)^{\frac{N}{M+1}}} \right\}^{M+1} \tag{11}$$

The MLE of the signal scaling and phase $\mu e^{j\alpha}$ for the probability density function under the alternate hypothesis H_1 is computed as follows:

$$\widehat{\mu e^{j\alpha}} = \frac{\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{y}}{\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{\Psi}} \tag{12}$$

We now have all the expressions of the maximum likelihood estimators of each parameter under each hypothesis and we can compute the generalized likelihood ratio for our binary hypothesis testing.

2.3.3 Expression of the GLR and link to the ASD test statistic

We can first start by computing the ratio of the two probability density functions detailed in Equation(11) and incorporating it the expression of $\widehat{\mu e^{j\alpha}}$, as the ratio between the two functions will only be the ratio of the noise scaling factor MLE. We eventually get an intermediate form of the GLR $\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y})$:

$$\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y}) = \left\{ \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right\}^N = \left\{ \frac{\underline{y}^H \boldsymbol{S}^{-1} \underline{y}}{(\underline{y} - \widehat{\mu e^{j\alpha}}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \widehat{\mu e^{j\alpha}}\underline{\Psi})} \right\}^N \tag{13}$$

The denominator of the expression in Equation(13) can then be simplified using developments of part 5.3:

$$(\underline{y} - \widehat{\mu e^{j\alpha}}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \widehat{\mu e^{j\alpha}}) = \underline{y}^H \boldsymbol{S}^{-1} \underline{y} - \frac{|\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{y}|^2}{\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{\Psi}}$$

Then dividing each term in the GLR by $\underline{y}^H \boldsymbol{S}^{-1} \underline{y}$ highlights the expression of our test statistic in the GLR and gives the final expression we needed to link our test statistic to the GLR :

$$\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y}) = \left\{ \frac{1}{1 - \frac{|\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{y}|^2}{(\underline{\Psi}^H \boldsymbol{S}^{-1} \underline{\Psi})(\underline{y}^H \boldsymbol{S}^{-1} \underline{y})}} \right\}^N$$

$$\hat{\mathcal{L}}(\boldsymbol{X}, \underline{y}) = \left\{ \frac{1}{1 - \widehat{\cos^2}} \right\}^N \tag{14}$$

The expression in Equation(14) is the major result of this paper, demonstrating that one can replace the GLR by the estimate of the cosine-squared in the GLRT for the binary hypothesis testing exposed in Equation(1) as the GLR is a monotone function of the estimate of the cosine-squared $\widehat{\cos^2}$.

3 Simulation Results

All the computations where made using `MatLab 2022b`.

A certain number of constants were arbitrarily fixed before the computations. The dimension N was fixed as $N = 10$. The noise structure R was arbitrarily generated as a random real symmetric and positive definite matrix in order to be able to generate synthetic training vectors and measurements. The steering vector $\underline{\Psi}$ was fixed as an arbitrarily generated complex vector.

Before studying the characteristics of the detector, we generated a set $\boldsymbol{X}$ of $M = 5000$ training vector $\underline{x_i} \sim \mathcal{CN}(\underline{0}, \boldsymbol{R})$ then we computed the covariance matrix of this set $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^H$.

3.1 Testing threshold η as a function of the probability of false alarm

We initially explored the relationship between the threshold of the GLRT (η) and the probability of false alarm (P_{fa}). The noise scaling factor σ^2 was set at a fixed value of 1. To achieve this, for a given P_{fa}, we generated $K = 200/P_{fa}$ measurement vectors to ensure a sufficient number of false alarms.

To experimentally determine the threshold, we first generated K measurements under the null hypothesis, representing a set of K vectors following $\mathcal{CN}(\underline{0}, \sigma^2\boldsymbol{R})$. For each measurement vector, we computed the test statistic $\widehat{\cos^2}$ using the defined expression in Equation (3), resulting in a vector of K values for the test statistic. This vector was then sorted in ascending order, and η was chosen to ensure the desired P_{fa}, i.e., selecting η such that its value ensures a number of false alarms equal to KP_{fa}. This was accomplished by choosing the value of the sorted test statistic vector located at the coordinate $\lfloor K * (1 - P_{fa}) \rfloor$. This procedure was repeated $MC = 100$ times to average the values obtained for each parameter using the Monte-Carlo method.

Varying P_{fa} affects both the number of generated measurement vectors and the proportion of false alarms considered. Figure (1) summarizes how the threshold η varies with P_{fa}. As expected, a higher P_{fa} corresponds to a lower threshold. Additionally, in line with the definition of cosine-squared, η ranges between 0 and 1, with higher P_{fa} values resulting in η closer to 0.

3.2 Probability of detection as a function of the Signal-to-Noise ratio (SNR)

As shown in our detection class, we used the following formula to compute the Signal-to-Noise Ratio (SNR) :

$$\text{SNR} = \mu^2 \underline{\Psi}^H (\sigma^2 \boldsymbol{R})^{-1} \underline{\Psi} = \mu^2 \underline{\Psi}^H \frac{1}{\sigma^2} \boldsymbol{R}^{-1} \underline{\Psi}$$

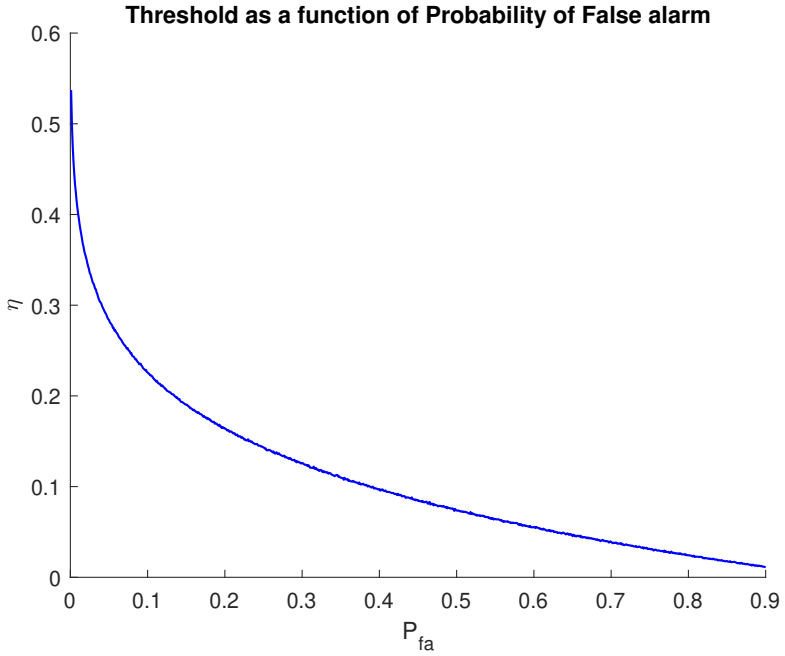

Figure 1: Plot of the value of the GLRT threshold as a function of the P_{fa}.

To study the evolution of the probability of detection P_d of this detector with respect to the SNR, one can vary only one parameter of the expression and keeping constant the others. We will choose to modify σ^2. We thus made two loop as we wanted to look at the evolution of P_d under different fixed P_{fa}.

For a fixed P_{fa} and iteration over σ^2—thus over SNR—we generated $K = 200/P_{fa}$ measurements under the null hypothesis $\underline{y}_{H_0} \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{R})$ to calculate the threshold η, following the same procedure as described in the preceding section (3.1). After this step, we generated $K = 200/P_{fa}$ measurements under the alternate hypothesis $\underline{y}_{H_1} \sim \mathcal{CN}(\mu e^{j\alpha} \underline{\Psi}, \sigma^2 \boldsymbol{R})$. Then we computed the test statistic $\widehat{\cos^2}$ of the measurements $\underline{y}_{H_1}$ and computed the number of values greater than η. This number was then divided by K to have an estimation of P_d. This procedure was done $MC = 100$ times to average the value we found for each parameter by Monte-Carlo method.

Figure(2) shows the results obtained with $\mu = 1$ under H_1, $\alpha = 0$, $\underline{\Psi} = (2 + 2i) \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$, the SNR was taken from 3 to 3.10^3 and three different values of $P_{fa} = 10^{-1}, 5.10^{-2}, 10^{-2}$. As expected, the higher the SNR, the higher P_d is. Also, for a fixed SNR, the higher P_fa the higher the probability of detection P_d will be, as we will have more distractors detected.

3.3 Receiving Operating Characteristics (ROC) Curve

Similar to our examination in Part (3.2), the ROC curve is constructed by calculating P_d as a function of P_{fa} and is frequently employed to assess the performance of a detector or classifier in computer science. It provides insights into the true positive rate versus the false positive rate.

Following a process close to what we did before, for a fixed SNR, we iteratively varied the values of P_{fa} and computed η for measurements under H_0. Subsequently, we determined P_d as described earlier for measurements under H_1. This process was repeated $MC = 100$ times to

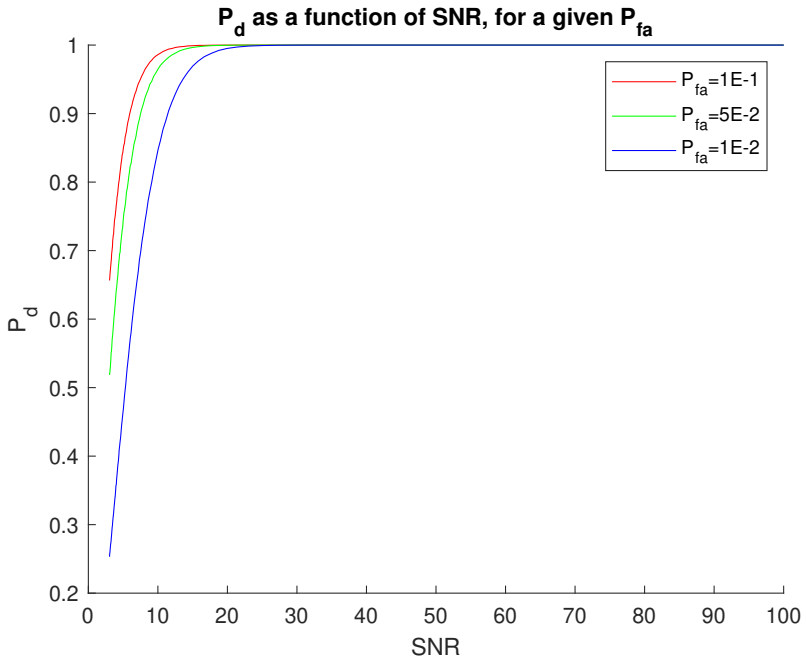

Figure 2: Plot of the value of P_d as a function of the SNR for different $P_{fa} = 10^{-1}, 5.10^{-2}, 10^{-2}, 10{-}3$.

obtain an average value for each parameter using the Monte-Carlo method.

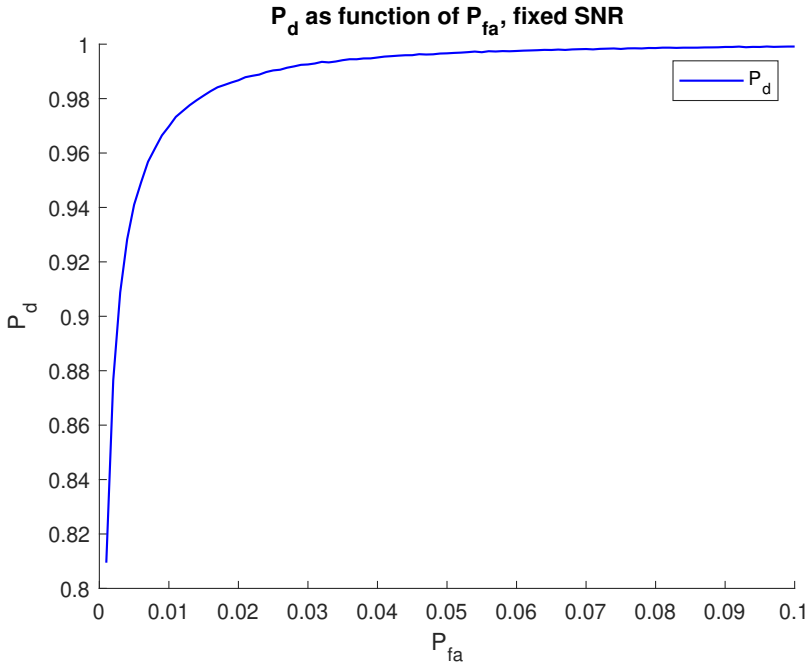

Figure 3: Plot of ROC curve for SNR= 15 and $P_{fa} = 10^{-3}$ to 10^{-1}.

Figure (3) illustrates the performance of the detector, emphasizing the trade-off between achieving a satisfactory P_d and maintaining a sufficiently low P_{fa}.

4 Synthesis and Conclusion

4.1 Summary of Findings

This article played a crucial role in expanding upon Kelly's paper [1], showcasing the development of an adaptive noise scaling invariant detector. The Generalized Likelihood Ratio Test (GLRT) introduced in this work demonstrated broader applicability compared to Kelly's original proposal. This discovery held significant importance at the time of the article's writing.

This report delves into the detailed computations, illustrating the process of establishing a connection between the proposed test statistic of the Adaptive Subspace Detector and the Generalized Likelihood Ratio through a monotone function, allowing to use ASD as a GLRT. This development facilitated the comparison of the test statistic to a threshold and enabled the assessment of the detector's performance through simulations on synthetic data.

4.2 Conclusion

This article exerted a substantial impact on the research community, accumulating nearly 500 citations at the time of composing this report. It stands as a significant achievement in the evolution of adaptive detection, building upon the groundwork laid by Kelly's prior work [1] published years earlier.

This report provides an academically rigorous yet concise exploration of the theory and concept involved in the article. It serves as an exemplary case for study and simulation using `MatLab`, highlighting various concepts within the detection domain.

For a more comprehensive evaluation, the applicability of this detector to real-world datasets - such as datasets from the Council of Scientific and Industrial Research - could offer a tangible understanding of its performance.

However, this article assumed the hypothesis of homogeneous training data, i.e., statistically independent and identically distributed, which is not perfectly met in real-life situations. This could result in a worse estimation of the sample covariance matrix $\boldsymbol{S}$. Several articles have addressed this phenomenon with various strategies. Among the latest articles [5] [7], a strategy relied on modeling this problem by representing training vectors as samples of different homogeneous clutter regions, each with different power. The goal was now to detect the clutter edges in the training vectors to get a better estimation of $\boldsymbol{S}$.

Many aspects still require attention in adaptive detection, but the article [2], which we studied in this report, remains one of the most important milestones in the history of detection theory, still relevant and utilized today.

5 Appendix

5.1 Computation of the MLE of the noise structure

This part aims to give detail to the computation of the MLE of the noise structure $\boldsymbol{R}$. he problem consisting in finding the estimate $\hat{\boldsymbol{R}}_{0,1}$ such that :

$$f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}) = \max_{\boldsymbol{R}} f_{0,1}(\boldsymbol{X}, \underline{y})$$

In order to maximize $f_{0,1}$ with respect to $\boldsymbol{R}$, we start from the expression in Equation(5) and we will maximize its logarithm, by computing the partial derivative with respect to $\boldsymbol{R}$ and finding its roots.

$$\frac{\partial \log(f_{0,1})}{\partial \boldsymbol{R}} = (M+1)\frac{\partial}{\partial \boldsymbol{R}} \log\left(\frac{1}{\pi^N ||\boldsymbol{R}|| \sigma^{\frac{2N}{M+1}}} \exp\left(-\text{tr}(\boldsymbol{R}^{-1}\boldsymbol{T}_{0,1})\right)\right)$$

Using basic logarithm properties concerning product and division to get rid of the constants, one can simplify the problem to the following equation :

$$\frac{\partial \log(f_{0,1})}{\partial \boldsymbol{R}} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial \boldsymbol{R}} \log(||\boldsymbol{R}||) + \frac{\partial}{\partial \boldsymbol{R}} \text{tr}(\boldsymbol{R}^{-1}\boldsymbol{T}_{0,1}) = 0$$

Using properties concerning derivatives of matrix operations with respect to a matrix detailed in Sections (2.1.4) and (2.5.4) of [3] and the fact that the noise structure is symmetric, we can simplify the following way :

$$\Rightarrow \quad \hat{\boldsymbol{R}}_{0,1}^{-1} + \hat{\boldsymbol{R}}_{0,1}^{-1}\boldsymbol{T}_{0,1}\hat{\boldsymbol{R}}_{0,1}^{-1} = 0$$

Then we easily find the expression of the Maximum Likelihood Estimate of the noise structure detailed in this report :

$$\Rightarrow \quad \hat{\boldsymbol{R}}_{0,1} = \boldsymbol{T}_{0,1}$$

5.2 Computation of the MLE of the noise scaling factor

One has to solve the following maximization problem to find the maximum likelihood estimate of the noise scaling factor σ^2 :

$$f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}, \hat{\sigma}^2) = \max_{\sigma^2} f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}) \tag{15}$$

The rewriting of the determinant of $\boldsymbol{T}_{0,1}$ following the process detailed in [1] after the injection of the expression of $\boldsymbol{R}_{0,1}$ involving the training vectors covariance matrix $\boldsymbol{S}$ allows us to work on a more convenient expression of the maximization problem :

$$f_{0,1}(\boldsymbol{X}, \underline{y}|\hat{\boldsymbol{R}}_{0,1}) = \left(\frac{1}{(e\pi)^N (\frac{M}{M+1})^N ||\boldsymbol{S}|| \sigma^{2\frac{N}{M+1}} \left(1 + \frac{1}{M\sigma^2}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \mu e^{j\alpha})\right)}\right)$$

This problem thus being the maximization of a ratio with the numerator being independent of σ^2, it is equivalent to the minimization of the denominator and we can also get rid of the multiplicative constants independent of σ^2, as following :

$$(17) \Leftrightarrow \min_{\sigma^2} \left(\sigma^{2\frac{N}{M+1}} \left(1 + \frac{1}{\sigma^2}(\underline{y} - \mu e^{j\alpha}\underline{\Psi})^H \boldsymbol{S}^{-1}(\underline{y} - \mu e^{j\alpha}\right)\right) \tag{16}$$

One can now compute the derivative of the function to be minimized in Equation(16) with respect to σ^2 and finding the roots of this derivative to find the maximum likelihood estimate of the noise scaling factor.

$$\frac{\partial}{\partial\sigma^2}\sigma^{2\frac{N}{M+1}}\left(1+\frac{1}{\sigma^2}(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})\right)=0$$

$$\Rightarrow \quad \frac{N}{M+1}\hat{\sigma}^{2\frac{N-M-1}{M+1}}+\frac{N-M-1}{M(M+1)}\hat{\sigma}^{2\frac{N-M-1}{M+1}}(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})=0$$

Knowing that the noise scaling factor is not null, we can divide each term by $\hat{\sigma}^{2\frac{N-M-1}{M+1}}$ and simplify by $\frac{1}{M+1}$ to get the following expression :

$$\Rightarrow \quad N+\frac{N-M-1}{M}(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})\frac{1}{\hat{\sigma}^2}=0$$

Isolating $\hat{\sigma}^2$ gives us the expression of the MLE of the noise scaling factor under each hypothesis :

$$\begin{cases}\hat{\sigma}_1^2=\frac{M-N+1}{MN}(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})\\ \hat{\sigma}_0^2=\hat{\sigma^2}_1|_{\mu=0}\end{cases}$$

5.3 Computation of the MLE of the signal scale and phase

One has to solve the following maximization problem to find the maximum likelihood estimate of the signal scale and phase $\mu e^{j\alpha}$:

$$f_1(\boldsymbol{X},\underline{y}|\hat{\boldsymbol{R}}_1,\hat{\sigma}^2,\widehat{\mu e^{j\alpha}})=\max_{\mu e^{j\alpha}} f_1(\boldsymbol{X},\underline{y}|\hat{\boldsymbol{R}}_1,\hat{\sigma}^2) \tag{17}$$

We will start from the maximization of the expression of Equation(11) with respect to $\mu e^{j\alpha}$:

$$f_1(\boldsymbol{X},\underline{y}|\hat{\boldsymbol{R}}_1,\hat{\sigma}_1^2)=\left\{\frac{(M+1)^N}{(e\pi M)^N||\boldsymbol{S}||(1+\frac{N}{M-N+1})\left((\frac{M-N+1}{MN})(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})\right)^{\frac{N}{M+1}}}\right\}^{M+1}$$

This maximization problem is equivalent to minimizing $(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})$ with respect to $\mu e^{j\alpha}$. This can be accomplished by completing the squares of the quadratic form, a process also mentionned in [1].

$$(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})=\underline{y}^H\boldsymbol{S}^{-1}\underline{y}+\mu^2\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}-2\mathrm{Re}(\mu e^{-j\alpha}\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{y})$$

As we want to get a single form involving $\mu e^{j\alpha}$ in this calculation, we will complete the square involving this term :

$$(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})=\underline{y}^H\boldsymbol{S}^{-1}\underline{y}+\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}\left\{|\mu e^{j\alpha}|^2-\mu e^{-j\alpha}\frac{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{y}}{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}}-\frac{\underline{y}^H\boldsymbol{S}^{-1}\underline{\Psi}}{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}}\mu e^{j\alpha}\right\}$$

$$(\underline{y}-\mu e^{j\alpha}\underline{\Psi})^H\boldsymbol{S}^{-1}(\underline{y}-\mu e^{j\alpha})=\underline{y}^H\boldsymbol{S}^{-1}\underline{y}+\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}\left|\mu e^{j\alpha}-\frac{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{y}}{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}}\right|^2-\frac{|\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{y}|^2}{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}}$$

We eventually get the expression of the MLE of the signal scale and phase :

$$\widehat{\mu e^{j\alpha}}=\frac{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{y}}{\underline{\Psi}^H\boldsymbol{S}^{-1}\underline{\Psi}}$$

5.4 GitHub repository for the MatLab code

You can find the code for this research on my GitHub: https://github.com/JulesGL/CFAR_ASD_Kraut.

References

[1] E.J. Kelly. "An Adaptive Detection Algorithm". In: *IEEE Transactions on Aerospace and Electronic Systems* AES-22.2 (1986), pp. 115–127. DOI: 10.1109/TAES.1986.310745.

[2] Shawn Kraut and Louis L. Scharf. "The CFAR adaptive subspace detector is a scale-invariant GLRT". In: *IEEE Transactions on Signal Processing* 47.9 (1998), pp. 2538–2541. DOI: 10.1109/78.782198.

[3] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Version 20121115. 2012. URL: http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html.

[4] L.L. Scharf and L.T. McWhorter. "Adaptive matched subspace detectors and adaptive coherence estimators". In: *Conference Record of The Thirtieth Asilomar Conference on Signals, Systems and Computers*. 1996, 1114–1117 vol.2. DOI: 10.1109/ACSSC.1996.599116.

[5] Bo Tang et al. "Adaptive Target Detection in Gaussian Clutter Edges". In: *IEEE Transactions on Aerospace and Electronic Systems* 56.2 (2020), pp. 1662–1673. DOI: 10.1109/TAES.2019.2930019.

[6] John M. Wozencraft and Irwin Mark Jacobs. *Principles of Communication Engineering by John M. Wozencraft and Irwin Jacobs*. Wiley, 1965.

[7] D. Xu et al. "Adaptive strategies for clutter edge detection in radar". In: *Signal Processing* 186 (2021), p. 108127. ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2021.108127. URL: https://www.sciencedirect.com/science/article/pii/S0165168421001651.