ADAPTIVE SUBSPACE DETECTOR IN HIGH DIMENSIONAL SPACE WITH INSUFFICIENT TRAINING DATA

Aref Miri Rekavandi, Abd-Krim Seghouane, Robin J. Evans

School of Electrical and Electronic Engineering
University of Melbourne, Parkville, VIC, Australia

ABSTRACT

Adaptive subspace detectors (ASD) generalize matched subspace detectors (MSD) by accounting for possible correlation. Both ASD and MSD are derived using the generalized likelihood ratio test (GLRT). While MSD assumes there is no correlation between observations, ASD estimates a sample covariance matrix of possibly correlated samples using signal-free observations. In this paper, we address the performance of the ASD when the number of secondary data is insufficient and the observed signal lies in higher dimensional space. Such high dimensional spaces are frequently encountered in functional magnetic resonance imaging (fMRI) data for the analysis of brain activation detection. We propose a methodology that works based on the latent variables in a lower dimensional space. A low-rank decomposition of the sample covariance matrix is derived based on the singular value decomposition (SVD) and an adaptive basis selection method is used to decide which eigen-vectors are useful in data projection. Performing detection in the lower dimensional subspace has the benefit of reducing the number of parameters which need to be estimated. Simulation results show superiority of our proposed adaptive reduced subspace detector (ARSD) over conventional ASD in term of probability of detection.

Index Terms— Detection, Likelihood Ratio Test, fMRI and Adaptive Reduced Subspace Detector.

1. INTRODUCTION

Signal detection in a noisy environment is an important problem that arises in many applications such as radar [1], communication [2] and medical imaging [3,4]. Neyman and Pearson [5] have shown that when parameters of the noise distribution are perfectly known, the likelihood ratio test is the uniformly most powerful (UMP) test which maximizes the probability of detection for a given false alarm rate. However, in many applications, the likelihood ratio test is unknown and therefore a UMP test cannot be derived. Therefore, several studies have investigated alternatives for the likelihood ratio test [6]. In [7] the matched subspace detector (MSD) was proposed based on the generalized likelihood ratio test (GLRT), for which all unknown parameters are replaced by their maximum likelihood values. In their study, [7] considers the noise samples to be independent and identically distributed (i.i.d) with a white Gaussian probability density. In another study, Kelly proposed a method that considers a covariance matrix for the noise distribution [8]. This covariance matrix is estimated from K signal-free samples. As an extension, [9] proposed a method that assumes both training data and test data have almost the same covariance structure, with only a scale factor difference. This method is also based on the likelihood ratio test and is named adaptive subspace detector (ASD). The ASD can be viewed as a generalized case of the MSD in which the data and subspaces are prewhitened using the sample covariance matrix. There are also alternative tests with different points of view, such as Rao test [10] and Wald test [11] or based on model selection criteria [12,13] that have been shown to be more suitable than the GLR under certain conditions. However, this study solely focuses on improving the performance of the GLRT.

The rest of the paper is organized as follows. In section 2 we present the background and problem statement. In section 3 we propose our methodology. In section 4 we present comparisons between conventional ASD and our proposed method. Finally, in section 5 we present some concluding remarks.

2. BACKGROUND

The detection problem that is addressed in our study is described as follows. We are given d samples from a real and scalar time series $\mathbf{y}(i), i = 0, 1, ..., d-1$ that is represented by the column vector $\mathbf{y}$. This vector of observations is generated by some components based on a general linear model (GLM):

$$\mathbf{y} = \mu \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{d \times p}$ is a known matrix whose columns span the signal subspace. It is also assumed that the columns of $\mathbf{H}$ are linearly independent. μ is a scalar factor that is non-zero when the signal is present. Entries of $\boldsymbol{\theta}$ contains unknown deterministic values. The noise components are $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{R})$, where σ^2 and $\mathbf{R}$ are unknown and should be estimated separately from training data.

The aim of detection is to decide between the null hypothesis $\mathcal{H}_0 : \mu = 0$ and the alternative hypothesis $\mathcal{H}_1 : \mu > 0$ for a given measurement vector $\mathbf{y}$. Given a set of K, i.i.d training noise vectors $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_K]$, each distributed as $\mathcal{N}(0, \mathbf{R})$, test signal vector $\mathbf{y} \in \mathbb{R}^d$ is measured and $y \sim \mathcal{N}(\mu\mathbf{H}\theta, \sigma^2\mathbf{R})$. It is clear that it is assumed $\sigma^2 = 1$ where $\mu = 0$. In [9] it was shown that when $\boldsymbol{\theta}$ and $\mathbf{R}$ are unknown, the following GLRT based test can be used to detect the signal presence.

$$l(\mathbf{y}) = \frac{\mathbf{y}^T\mathbf{S}^{-\frac{1}{2}}\mathbf{P_D}\mathbf{S}^{-\frac{1}{2}}\mathbf{y}}{\mathbf{y}^T\mathbf{S}^{-1}\mathbf{y}} \tag{2}$$

where $\mathbf{D}$ is $\mathbf{S}^{-\frac{1}{2}}\mathbf{H}$ and $\mathbf{P_D} = \mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T$ is the projection operator that projects the whitened observation vector $\mathbf{y}$ onto the subspace that is spanned by columns of $\mathbf{D}$. The matrix

$$\mathbf{S} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{n}_i\mathbf{n}_i^T \tag{3}$$

is the sample covariance matrix, which is estimated from the K training data. This matrix $\mathbf{S}$ is actually an estimation of covariance matrix of noise $\mathbf{R}$, that maximizes likelihood function. This solution for signal detection is referred as ASD.

In general, for signal detection using (1) estimation of $p + d(d+1)/2$ unknown parameters is required where p is the number of entries of vector $\boldsymbol{\theta}$ and d is dimension of the symmetric covariance matrix $\mathbf{R}$. Therefore, the number of parameters to be estimated is of order $O(d^2)$. For a limited training data and a high dimensional measurement vector $\mathbf{y}$, (in some applications such as radar and medical imaging) applying ASD directly on data is inefficient and parameters are highly deviated from their actual values. Therefore, introducing knowledge about the covariance matrix is required to improve the performance of ASD.

3. PROPOSED METHOD

As mentioned earlier, in high dimensional space hypothesis testing, there are many unknown parameters (in the order of $O(d^2)$) that should be estimated. Several studies have addressed this issue by assuming a known structure for the covariance matrix [14]. One of the ways of introducing structure to the covariance matrix in order to improve performance of the detector (2), is to assume that the measurement vector $\mathbf{y}$ and noise vector $\mathbf{n}$ depend on latent variables in a lower dimensional space. In other words, when the number of training data is insufficient, the covariance matrix can be described by a few dominant eigen-vectors/values. Based on this assumption, there are two latent variables $\mathbf{z} \in \mathbb{R}^{d_1}$ and $\mathbf{e} \in \mathbb{R}^{d_1}$ with $d_1 \ll d$ and

$$\mathbf{y} = \mathbf{U}_{d_1}\mathbf{z} \quad \text{and} \quad \mathbf{n} = \mathbf{U}_{d_1}\mathbf{e} \tag{4}$$

where $\mathbf{U}_{d_1}$ is an unknown $d \times d_1$ matrix whose columns are orthogonal, (i.e. $\mathbf{U}_{d_1}^T\mathbf{U}_{d_1} = \mathbf{I}_{d_1}$, where $\mathbf{I}_{d_1}$ is the identity matrix of size $d_1 \times d_1$). It is also assumed that $\{\mathbf{e}\}$ is a zero mean Gaussian noise vector with a $d_1 \times d_1$ covariance matrix $\boldsymbol{\Lambda}$. Substituting (4) in (1) we have

$$\mathbf{U}_{d_1}\mathbf{z} = \mu\mathbf{H}\boldsymbol{\theta} + \mathbf{U}_{d_1}\mathbf{e} \tag{5}$$

Multiplying both sides by $\mathbf{U}_{d1}^T$ gives

$$\mathbf{z} = \mu\mathbf{U}_{d_1}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{e} \tag{6}$$

In lower dimensional space, the formulation of (6) is the same as (1) and we can apply standard detection theory results on the latent variable $\mathbf{z}$ with the difference that the signal subspace in the lower dimensional space is the span of the columns of $\mathbf{U}_{d_1}^T\mathbf{H}$, instead of $\mathbf{H}$. Therefore, the number of unknown parameters is reduced to $p + d_1(d_1+1)/2$ which is significantly less than $p + d(d+1)/2$, corresponding to the higher dimensional space. In general, using an optimum projection for the lower dimensional space, we may lose some useful information, however, the advantage is that we only need to estimate a small number of parameters. Reducing the number of parameters reduces the parameter estimation error too.

The ASD for the latent variable $\mathbf{z}$ is:

$$l(\mathbf{z}) = \frac{\mathbf{z}^T\hat{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{P_B}\hat{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{z}}{\mathbf{z}^T\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{z}} \tag{7}$$

where $\mathbf{B}$ is $\hat{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{U}_{d_1}^T\mathbf{H}$ and

$$\hat{\boldsymbol{\Lambda}} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{U}_{d_1}^T\mathbf{n}_i\mathbf{n}_i^T\mathbf{U}_{d_1} = \mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1} \tag{8}$$

Finally, using threshold value τ, we can decide between $\mathcal{H}_0$ and $\mathcal{H}_1$. We name our reduced version of ASD, the adaptive reduced subspace detector (ARSD).

3.1. Estimation of $\mathbf{U}_{d_1}$

The primary challenge of ARSD is finding the best projection matrix $\mathbf{U}_{d_1}$. When the number of training data is insufficient, the sample covariance matrix is not consistent and the eigenvalues and eigen-vectors of the sample covariance matrix can be significantly different from their true values [15].

In order to find a suitable projection basis set, we need to investigate equation (7) in detail. After reparameterizing of the reduced version of the test in terms of $\mathbf{y}$, we will have

$$l'(\mathbf{y}) = \frac{\mathbf{y}^T\mathbf{U}_{d_1}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{U}_{d_1}^T\mathbf{H}(\mathbf{H}^T\mathbf{U}_{d_1}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{U}_{d_1}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{U}_{d_1}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{U}_{d_1}^T\mathbf{y}}{\mathbf{y}^T\mathbf{U}_{d_1}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{U}_{d_1}^T\mathbf{y}} \tag{9}$$

Algorithm 1 Proposed ARSD

Input: Observation vector $\mathbf{y}$, Training data $\mathbf{N}$, $\mathbf{H}$, d_1,α, τ
Output: Decision on $\mathcal{H}_0$ or $\mathcal{H}_1$

procedure ARSD
  $\mathbf{S} \leftarrow \frac{1}{K}\sum_{i=1}^{K} \mathbf{n}_i\mathbf{n}_i^T$
  $\mathbf{U}, \Gamma, \mathbf{U}^T \leftarrow SVD(\mathbf{S})$
  $\mathbf{U}_{d_1} \leftarrow$ solve equation (11) or (16)
  $\mathbf{z} \leftarrow \mathbf{U}_{d_1}^T\mathbf{y}$
  $\hat{\mathbf{\Lambda}} \leftarrow \frac{1}{K}\sum_{i=1}^{K} \mathbf{U}_{d_1}^T\mathbf{n}_i\mathbf{n}_i^T\mathbf{U}_{d_1}$
  $\mathbf{B} \leftarrow \hat{\mathbf{\Lambda}}^{-\frac{1}{2}}\mathbf{U}_{d_1}^T\mathbf{H}$
  $l(\mathbf{z}) \leftarrow \frac{\mathbf{z}^T\hat{\mathbf{\Lambda}}^{-\frac{1}{2}}\mathbf{P}_B\hat{\mathbf{\Lambda}}^{-\frac{1}{2}}\mathbf{z}}{\mathbf{z}^T\hat{\mathbf{\Lambda}}^{-1}\mathbf{z}}$
  Output $\leftarrow \mathcal{H}_0$
  if $l(\mathbf{z}) \geq \tau$ **then**
    Output $\leftarrow \mathcal{H}_1$

In comparison with the test shown in equation (2), the only difference is the estimated covariance matrix which is a modified form of the sample covariance matrix, that is:

$$\hat{\mathbf{R}}_{Mod}^{-1} = \mathbf{U}_{d_1}\hat{\Lambda}^{-1}\mathbf{U}_{d_1}^T = \mathbf{U}_{d_1}(\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1})^{-1}\mathbf{U}_{d_1}^T \tag{10}$$

We would like to find a basis set $\mathbf{U}_{d_1}$ that improves the overall performance of the detector in (7). For finding the best projection matrix, we consider two different cases of covariance structures on which our method can be applied.

3.1.1. *$\mathbf{R}$ with uniform diagonal components*

When the covariance matrix has the same diagonal components, it means that the variance of the noise is fixed for all components of noise vector $\mathbf{n}$, which can happen in many applications. In this case study, we need to solve the following optimization problem:

$$\hat{\mathbf{U}}_{d_1} = \arg\min_{\mathbf{U}_{d_1}} \|\mathbf{U}_{d_1}\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}\mathbf{U}_{d_1}^T - \mathbf{S}\|_F^2 \tag{11}$$

where $\|.\|_F^2$ shows Frobenius norm of a matrix. This optimization problem means that we need to regularize the sample covariance in a way that the result gets close enough to the sample covariance. For a known d_1 the solution is the first d_1 eigen-vectors of $\mathbf{S}$.

3.1.2. *Ill-conditioned $\mathbf{R}$*

In the previous section, the solution was the first eigen-vectors of the sample covariance. It means that the last eigen-vectors were in the direction of the noise which is caused by insufficiency in the training data. In this new case, when the covariance matrix is in ill condition, we cannot have the same strategy. In this case, $\mathbf{R}^{-1}$ is significantly dependent on the last eigen-vectors and we should be careful about picking eigen-vectors. A standard approach is to find the maximum likelihood (ML) estimate for $\mathbf{U}_{d_1}$. To compute the likelihood for the given projected training data we have:

$$f[\mathbf{U}_{d_1}^T\mathbf{n}_1, ..., \mathbf{U}_{d_1}^T\mathbf{n}_K] = \Big[\frac{1}{\pi^{d1}|\mathbf{U}_{d_1}^T\mathbf{R}\mathbf{U}_{d_1}|}\exp\{-tr((\mathbf{U}_{d_1}^T\mathbf{R}\mathbf{U}_{d_1})^{-1}\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1})\Big]^K \tag{12}$$

where $|.|$ shows the determinant of a matrix. This likelihood function has two variables $\mathbf{R}$ and $\mathbf{U}_{d_1}$ that need to be estimated via ML. To do this, we follow a two-stage estimation procedure. We first assume $\mathbf{U}_{d_1}$ is fixed and estimate $\mathbf{R}$. In the second stage, we keep $\mathbf{R}$ fixed and estimate $\mathbf{U}_{d_1}$. Assuming $\mathbf{U}_{d_1}$ is fixed, we have:

$$\mathbf{U}_{d_1}^T\mathbf{R}\mathbf{U}_{d_1} = \mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1} \tag{13}$$

The optimal $\mathbf{R}$ must satisfy (13). Therefore, we now replace $\mathbf{U}_{d_1}^T\mathbf{R}\mathbf{U}_{d_1}$ in (12) with $\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}$ and after simplification

$$f[\mathbf{U}_{d_1}^T\mathbf{n}_1, ..., \mathbf{U}_{d_1}^T\mathbf{n}_K] \propto \frac{1}{|\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}|} \tag{14}$$

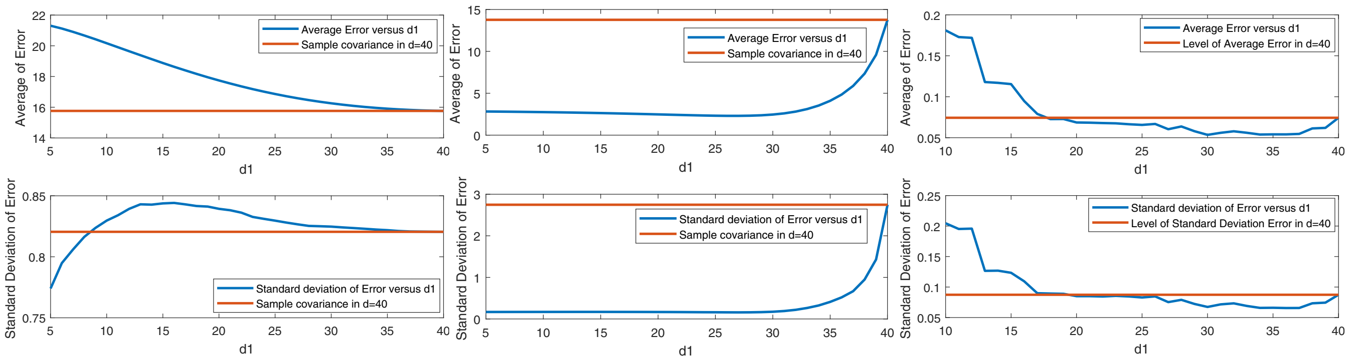
Now, maximizing the likelihood with respect to $\mathbf{U}_{d_1}$ can be changed to the following problem

$$\hat{\mathbf{U}}_{d_1} = \arg\min_{\mathbf{U}_{d_1}} |\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}| \tag{15}$$

In (15), if d_1 is assumed to be known, the solution is the last d_1 eigen-vectors which can be obtained by singular value decomposition of the sample covariance matrix $\mathbf{S}$. So far, we have found that to maximize the likelihood function, we need to pick the last eigen-vectors. However, we know that although $\mathbf{S}$ is highly deviated from the actual covariance when the number of training samples is insufficient, but the sample covariance is a consistent estimation when data is enough. Therefore,our solution must also be as close as possible to the sample covariance matrix $\mathbf{S}$ in the Frobenius norm sense (similar to 11). As a result, instead of a conventional maximum likelihood solution for $\mathbf{U}_{d_1}$, we are faced with a constrained optimization problem which can be written as a multi objective optimization problem:

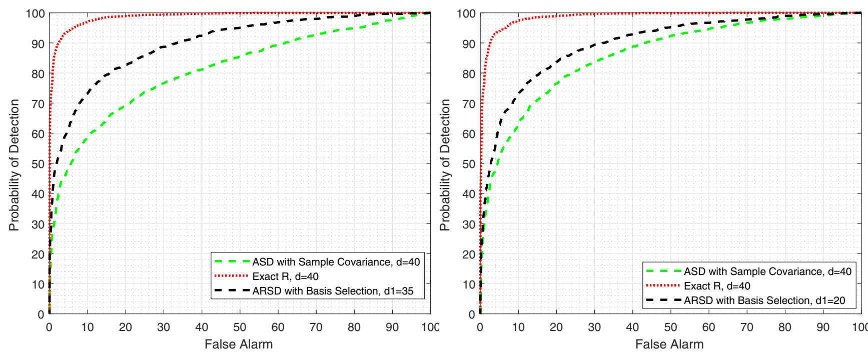
$$\hat{\mathbf{U}}_{d_1} = \arg\min_{\mathbf{U}_{d_1}} |\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}| + \alpha\|\mathbf{U}_{d_1}\mathbf{U}_{d_1}^T\mathbf{S}\mathbf{U}_{d_1}\mathbf{U}_{d_1}^T - \mathbf{S}\|_F^2 \tag{16}$$

which tries to maximize the likelihood such that the estimated covariance remains in the neighborhood of the sample covariance. The eigen-vectors that keep the covariance as close as possible to the sample covariance in terms of Frobenius norm, are the first eigenvectors of $\mathbf{S}$. Meanwhile, the last eigen-vectors minimize the determinant, the first term of (16). Finally a combination of the first a and the last b eigen-vectors of $\mathbf{S}$ is the solution of problem (16) where $a + b = d_1$. Algorithm 1 summarizes the basis selection method for ARSD.



(a) $\hat{\mathbf{R}}_{Mod}$ (b) $\hat{\mathbf{R}}_{Mod}^{-1}$ (c) $\hat{\boldsymbol{\theta}}$

Fig. 1: Average and standard deviation of Frobenius norm of difference between the actual and the estimated (a) covariance, (b) covariance inverse and (c) $\boldsymbol{\theta}$.



(a) (b)

Fig. 2: ROC of detectors in SNR=12 dB (a) $\mathbf{R}$ with uniform diagonal components, (b) Ill-conditioned $\mathbf{R}$

4. SIMULATION RESULTS

In this section, we address the performance of ARSD in comparison with conventional ASD when the training data is insufficient. During simulation, we assume $d = 40$, and the number of training data K is 60, which means that K is not significantly larger than d. In lower dimensional space, we can change the dimension to the d_1 that makes the K/d_1 large. We randomly produce the signal subspace matrix $\mathbf{H}$ of size 40×2 (based on event related stimulus function of fMRI data analysis [16, 17]) and it is fixed for all simulations. First, using equations (8), (11) and (16), parameters $\boldsymbol{\Lambda}$ and $\mathbf{U}_{d_1}$ are estimated in lower dimensional space. These parameters are then used in ARSD (7) and will assess the probability of detection. Figure 1 shows the average and standard deviation of Frobenius norm of error between the estimated and actual $\boldsymbol{\theta}$, $\mathbf{R}$ and $\mathbf{R}^{-1}$ where a random uniform covariance is used. It can be seen that when we bring the observation vector to a lower dimensional space for a range of d_1, we can have a better estimation of $\boldsymbol{\theta}$ which makes our test performance better (see figure 1c).
Finally figure 2 illustrates the performance of the proposed detector in comparison with conventional ASD using a receiver operating characteristic (ROC) curve. In this figure, the performance of the proposed method is shown where we can see that ARSD (black) always outperforms ASD (green). The red ROC curve shows the oracle scenario where the covariance matrix is known. When we increase the value K, both of ASD and ARSD plots converge to the red plot.

5. CONCLUSION

In this study, a reduced version of the adaptive subspace detector (ASD) was proposed, namely adaptive reduced subspace detector (ARSD). It was shown for some scenarios, by bringing data to a lower dimensional space, the likelihood ratio test can provide a better probability of detection. The most important section of the algorithm is providing proper orthonormal basis to project data. We have used a basis selection method that picks some eigen-vectors among the sorted eigenvectors of the sample covariance to minimize our defined objective function.

6. REFERENCES

[1] Wai-Sheou Chen and Irving S Reed, "A new cfar detection test for radar," *Digital Signal Processing*, vol. 1, no. 4, pp. 198–214, 1991.

[2] Michael L McCloud and Louis L Scharf, "Interference estimation with applications to blind multiple-access communication over fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 947–961, 2000.

[3] A. K. Seghouane, "fMRI activation detection using a variant of Akaike Information Criterion," *In Proceedings of the IEEE Workshop on Statistical Signal Processing, SSP*, pp. 233–236, 2012.

[4] A. K. Seghouane and J. L. Ong, "A Bayesian model selection approach to fMRI activation detection," *In Proceedings of IEEE International Conference on Image Processing, ICIP*, pp. 4401–4404, 2010.

[5] Jerzy Neyman and Egon S Pearson, "IX. on the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, vol. 231, no. 694-706, pp. 289–337, 1933.

[6] Ali Ghobadzadeh, Saeed Gazor, Masoud Naderpour, and Ali Akbar Tadaion, "Asymptotically optimal cfar detectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 897–909, 2016.

[7] Louis L Scharf and Benjamin Friedlander, "Matched subspace detectors," *IEEE Transactions on signal processing*, vol. 42, no. 8, pp. 2146–2157, 1994.

[8] Edward J Kelly, "An adaptive detection algorithm," *IEEE transactions on aerospace and electronic systems*, , no. 2, pp. 115–127, 1986.

[9] Shawn Kraut, Louis L Scharf, and L Todd McWhorter, "Adaptive subspace detectors," *IEEE Transactions on signal processing*, vol. 49, no. 1, pp. 1–16, 2001.

[10] Antonio De Maio, "Rao test for adaptive detection in gaussian interference with unknown covariance matrix," *IEEE transactions on signal processing*, vol. 55, no. 7, pp. 3577–3584, 2007.

[11] Antonio De Maio, "A new derivation of the adaptive matched filter," *IEEE Signal Processing Letters*, vol. 11, no. 10, pp. 792–793, 2004.

[12] A. K. Seghouane, "Asymptotic bootstrap corrections of AIC for linear regression models," *Signal Processing*, vol. 90, pp. 217–224, 2010.

[13] A. K. Seghouane, "New AIC corrected variants for multivariate linear regression model selection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 1154–1165, 2011.

[14] Satyabrata Sen, "Low-rank matrix decomposition and spatio-temporal sparse recovery for STAP radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1510–1523, 2015.

[15] Iain M Johnstone and Arthur Yu Lu, "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 682–693, 2009.

[16] A. K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Transactions on Medical Imaging*, pp. 1796–1807, 2017.

[17] A. K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Transactions on Image Processing*, pp. 3002–3015, 2017.