A SUBSPACE APPROACH FOR SHRINKAGE PARAMETER SELECTION IN UNDERSAMPLED CONFIGURATION FOR REGULARISED TYLER ESTIMATORS

Q. Hoarau[1], *A. Breloy*[2], *G. Ginolhac*[1], *A.M. Atto*[1], *J.M. Nicolas*[3]

[1] LISTIC, Université Savoie Mont Blanc, France
[2] LEME, Université Paris X, France
[3] LTCI, Télécom ParisTech, France

ABSTRACT

Regularized Tyler Estimator's (RTE) have raised attention over the past years due to their attractive performance over a wide range of noise distributions and their natural robustness to outliers. Developing adaptive methods for the selection of the regularisation parameter α is currently an active topic of research. Indeed, the bias-performance compromise of RTEs highly depends on the considered application. Thus, finding a generic rule that is optimal for every criterion and/or data configurations is not straightforward. This issue is addressed in this paper for undersampled configurations (number of samples lower than the dimension of the data). The paper proposes a new regularisation parameter selection based on a subspace reduction approach. The performance of this method is investigated in terms of estimation accuracy and for adaptive detection purposes, both on simulation and real data.

***Index Terms*—** regularised covariance matrix estimation, robust estimation, adaptive detection, subspace.

1. INTRODUCTION

Covariance matrix estimation is a key step for many applications in signal processing. In array processing, for instance, the estimation accuracy of this parameter has a direct impact on the performance of adaptive detectors. Given a set of K samples $\{\boldsymbol{x}_k\} \in \mathbb{C}^N$, the Sample Covariance Matrix (SCM) is the most frequently used covariance matrix estimator. However, it is known to be a poor covariance matrix estimator when samples are drawn from heavy tailed distributions and/or corrupted by outliers.

To overcome this issue, the M-Estimators (generalized maximum likelihood estimators on the class of CES distributions) have recently attracted considerable interest due to their robustness properties [1]. Nevertheless, these estimators are not suited to the now common problem of high dimensional data with low sample support. More specifically M-estimators are not defined in undersampled configuration ($K < N$) and the typical rule of thumb suggest that $K > 2N$ is required in order to reach good performance.

Authors would like to thank BPI France for funding the project, and ENGIE for providing real datasets.

This is why shrinkage (or regularisation) methods have been recently proposed to deal with these issues [2, 3, 4, 5, 6]. Currently, the regularised Tyler estimators (RTEs), which are expressed as:

$$\hat{\boldsymbol{R}}_{RTE}(\alpha) = (1-\alpha)\frac{N}{K}\sum_{k=1}^{K}\frac{\boldsymbol{x}_k\,\boldsymbol{x}_k^H}{\boldsymbol{x}_k^H\,\hat{\boldsymbol{R}}_{RTE}^{-1}(\alpha)\boldsymbol{x}_k} + \alpha\boldsymbol{I}_N \quad (1)$$

are especially receiving increasing interest [7, 8, 9, 10, 11, 12, 13]. These estimators correspond to regularised versions of the robust Tyler estimator, where the shrinkage towards the identity matrix ensures existence of the estimator in undersampled configuration (RTEs exist and are unique for $\alpha \in (max(0, 1-K/N); 1])$, as well a good conditioning. Moreover estimators can be computed using the fixed point iterations:

$$\hat{\boldsymbol{R}}^{(i+1)} = (1-\alpha)\frac{N}{K}\sum_{k=1}^{K}\frac{\boldsymbol{x}_k\,\boldsymbol{x}_k^H}{\boldsymbol{x}_k^H\,\hat{\boldsymbol{R}}^{(i)\,-1}\boldsymbol{x}_k} + \alpha\boldsymbol{I}_N\,, \quad (2)$$

that converge to the solution $\hat{\boldsymbol{R}}_{RTE}(\alpha)$.

With this new class of estimators, arises the highly nontrivial issue of the selection of the regularisation parameter α. Indeed, the bias-performance compromise of RTEs highly depends on the considered application, so one can not expect a generic rule that is optimal for every criterion and/or data configurations. From the recent state of the art, one can classify several approaches:

- Oracle scheme associated to algorithm (2): [5] minimises the expected error for shape (scaling-free) estimation.
- Random Matrix Theory (RMT) regime (i.e. both K and N tends to infinity at fixed rate K/N) estimators : [7] minimises the MSE, [9] ensures optimal performance of the ANMF detector, [8] minimises portfolio variance (which similar to inverse SINR-Loss).
- Alternate variants of the algorithm (2) (e.g. trace normalized iterations): [2] oracle estimator that minimises the MSE, [14, 15] maximises the expected likelihood ratio, [13] proposes a modified normalisation.

Regarding to this range of methods, we propose in this paper to adapt oracle schemes of [2, 5] for undersampled configurations ($K < N$) using a dimension reduction approach [15]. Our motivations are are twofold:

- Oracle schemes have closed form expressions and are not requiring an expansive grid search as in e.g. [9]. Hence they are more suitable to implementation.
- The current oracle schemes proposed in [2, 5] are not well suited for severely undersampled configurations, as it will be explained and illustrated in the simulation section of this paper.

As a by product, we also re-derive oracle estimators of [2] for the algorithm (2), both for the real and complex cases. The performance of this method is investigated in terms of estimation accuracy and for adaptive detection purposes, both on simulation and real data.

2. CONTEXT

2.1. Regularised Tyler Estimator

Let $\{\boldsymbol{x}_k\}_{1,K} \in \mathbb{C}^N$ a set of K N-variate CES distributed random vectors $\boldsymbol{x}_k \sim CE_N(\mathbf{0}, \boldsymbol{R}, g)$, of mean $\mathbf{0} \in \{0\}^N$, scatter matrix $\boldsymbol{R} \in \mathbb{C}^{N\times N}$ and generator g. The maximum likelihood estimator of $\boldsymbol{R}$ minimises the negative log-likelihood:

$$\mathcal{L}(\boldsymbol{R}) = \frac{1}{K}\sum_{k=1}^{K} \rho(\boldsymbol{x}_k^H \boldsymbol{R}^{-1}\boldsymbol{x}_k) - \ln |\boldsymbol{R}^{-1}| \,, \tag{3}$$

with $\rho(t) = -\ln g(t)$. This MLE can be generalised to obtain M-Estimators of the scatter matrix by choosing more general ρ functions. In the particular case where $\rho(t) = N \ln t$, (3) leads to the Tyler Estimator [16]:

$$\hat{\boldsymbol{R}}_{TE} = \frac{N}{K}\sum_{k=1}^{K} \frac{\boldsymbol{x}_k\boldsymbol{x}_k^H}{\boldsymbol{x}_k^H \hat{\boldsymbol{R}}_{TE}^{-1}\boldsymbol{x}_k} \,. \tag{4}$$

When the sample support K is not sufficient ($K < 2N$), the matrix $\hat{\boldsymbol{R}}_{TE}$ is ill-conditioned, and for $K < N$ it does not exist. The estimator is thus regularised in order to solve the issue, this operation being also known as 'shrinkage towards identity.' A first algorithm for the regularised Tyler estimator (RTE) has been proposed in [17]

$$\begin{cases} \tilde{\boldsymbol{R}}_{CWH}^{(i+1)} = (1-\alpha)\dfrac{N}{K}\displaystyle\sum_{k=1}^{K} \dfrac{\boldsymbol{x}_k\,\boldsymbol{x}_k^H}{\boldsymbol{x}_k^H \hat{\boldsymbol{R}}_{CWH}^{(i)\,-1}\boldsymbol{x}_k} + \alpha\,\boldsymbol{I}_N, \\ \hat{\boldsymbol{R}}_{CWH}^{(i+1)} = N\dfrac{\tilde{\boldsymbol{R}}_{CWH}^{(i+1)}}{\mathrm{Tr}(\tilde{\boldsymbol{R}}_{CWH}^{(i+1)})}. \end{cases} \tag{5}$$

These fixed point iterations converge for any $\alpha \in (0;1)$ as proven in [2], however this is a heuristic solution as it does not follows from the minimisation of an objective function. The RTE have been proposed as minimiser of a penalised negative log-likelihood in [4, 5, 6]

$$\mathcal{L}_\alpha(\boldsymbol{R}) = \frac{1}{K}\sum_{k=1}^{K} \rho(\boldsymbol{x}_k^H \boldsymbol{R}^{-1}\boldsymbol{x}_k) - \ln |\boldsymbol{R}^{-1}| + \alpha\mathcal{P}(\boldsymbol{R}) \,, \tag{6}$$

with $\mathcal{P}(\boldsymbol{R}) = \mathrm{Tr}(\boldsymbol{R}^{-1})$. This yields the unique solution in (1) that can be computed using the iterations (2), which converges for $\alpha \in (max(0, 1 - \frac{K}{N}); 1]$. We will focus on this algorithm for the rest of the paper.

2.2. Oracle shrinkage parameter selection

Depending on the application, the parameter α must be chosen wisely. [2] proposed an oracle by minimising the MSE:

$$\alpha_{CWH} = \underset{\alpha}{\mathrm{argmax}}\; E[\|\boldsymbol{R} - \hat{\boldsymbol{R}}\|^2] \,, \tag{7}$$

Proposition 2.1 *Let* $\mathfrak{C}_1 = ((N+1)(K+1) - 4)\, Tr(\boldsymbol{R}^2)$ *and* $\mathfrak{C}_2 = (KN + K - 1)\, Tr(\boldsymbol{R}^2)$*. For i.i.d CES distributed samples, the solution to (7) when using algorithm (2) is, in the real case:*

$$\alpha_{CWH} = \frac{(N-2)\, Tr(\boldsymbol{R}^2) + N\, Tr^2(\boldsymbol{R})}{\mathfrak{C}_1 + N\, Tr^2(\boldsymbol{R}) - 2K(N+2)\, Tr(\boldsymbol{R}) + KN(N+2)}, \tag{8}$$

and in the complex case:

$$\alpha_{CWH} = \frac{N\, Tr^2(\boldsymbol{R}) - Tr(\boldsymbol{R}^2)}{\mathfrak{C}_2 + N\, Tr^2(\boldsymbol{R}) - 2K(N+1)\, Tr(\boldsymbol{R}) + KN(N+1)} \tag{9}$$

Proof: See [2], with minor adaptation for the complex case and, since we work with algorithm (2), the normalisation $\mathrm{Tr}(\boldsymbol{R}) = N$ cannot be applied.

Another criterion was proposed in [5], focused on fitting the shape of the matrix rather than its values:

$$\alpha_{OTy} = \underset{\alpha}{\mathrm{argmax}}\; E[\|\boldsymbol{R}^{-1}\hat{\boldsymbol{R}} - \frac{\mathrm{Tr}(\boldsymbol{R}^{-1}\hat{\boldsymbol{R}})}{N}\,\boldsymbol{I}_N\|^2] \,, \tag{10}$$

Proposition 2.2 *For i.i.d CES distributed samples, the solution to (10) when using algorithm (2) is, in the real case:*

$$\alpha_{OTy} = \frac{N - 2 + N\, Tr(\boldsymbol{R})}{N - 2 + N\, Tr(\boldsymbol{R}) + K(N+2)\left(\frac{Tr(\boldsymbol{R}^{-2})}{N} - 1\right)} \tag{11}$$

and in the complex case:

$$\alpha_{OTy} = \frac{N\, Tr(\boldsymbol{R}) - 1}{N\, Tr(\boldsymbol{R}) - 1 + K(N+1)\left(\frac{Tr(\boldsymbol{R}^{-2})}{N} - 1\right)} \,. \tag{12}$$

Note that these α are oracles as they require to know the values of $\text{Tr}(\boldsymbol{R})$ and $\text{Tr}(\boldsymbol{R}^2)$. The true covariance matrix value being unknown, it can be replaced by a RTE as proposed in [5]. In the case were $K \simeq N$, this estimator seems, in practice, accurate enough. However, when $K \ll N$ the traces estimation becomes greatly inaccurate, enforcing $\hat{\alpha}_{OTy} \simeq 1$ and impeding the regularisation process or causing $\hat{\alpha}_{CWH} < 1 - \frac{K}{N}$ and the divergence of the recursive algorithm. Aiming to solve those issues, we propose a new scheme for the estimation of covariance matrices in the undersampled configuration, inspired by the work of [15].

3. SUBSPACE APPROACH

The estimation procedure we propose consists in (i) projecting the samples onto the K-subspace they span, (ii) estimating covariance matrix onto this subspace using a RTE and (iii) expanding this result to a $N \times N$ estimator.

3.1. Dimension reduction

The first step is to retrieve the subspace spanned by the sample vectors $\{\boldsymbol{x}_k\}$ by performing the SVD of the SCM:

$$\hat{\boldsymbol{R}}_{SCM} = \hat{\boldsymbol{U}}\,\hat{\boldsymbol{D}}\,\hat{\boldsymbol{U}}^H, \tag{13}$$

with $\hat{\boldsymbol{D}}$ a diagonal matrix and $\hat{\boldsymbol{U}} = [\hat{\boldsymbol{u}}_1 \dots \hat{\boldsymbol{u}}_N]$. In the case $K < N$, only K eigenvalues are non zero and their corresponding eigenvectors $\hat{\boldsymbol{U}}_K = [\hat{\boldsymbol{u}}_1 \dots \hat{\boldsymbol{u}}_K]$ (first K columns of $\hat{\boldsymbol{U}}$) span the subspace where the data lies in. Also denote $\hat{\boldsymbol{U}}_K^{\perp} = [\hat{\boldsymbol{u}}_{K+1} \dots \hat{\boldsymbol{u}}_N]$. Using this basis, the true covariance matrix $\boldsymbol{R}$ is decomposed in blocks as:

$$\boldsymbol{R} = \left[\hat{\boldsymbol{U}}_K\,\hat{\boldsymbol{U}}_K^{\perp}\right] \begin{bmatrix} \boldsymbol{R}_K & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_K^{\perp} \end{bmatrix} \left[\hat{\boldsymbol{U}}_K\,\hat{\boldsymbol{U}}_K^{\perp}\right]^H. \tag{14}$$

As argued in [15] inference about the covariance matrix is possible only in the K-dimensional subspace spanned by the data. Hence one can only estimate $\boldsymbol{R}_K$, which can be done using the projections of the vectors $\boldsymbol{x}_k$ onto this subspace:

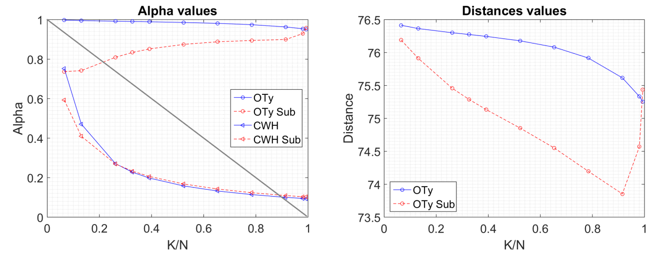
$$\tilde{\boldsymbol{x}}_k = \hat{\boldsymbol{U}}_K^H\,\boldsymbol{x}_k\,. \tag{15}$$

3.2. RTE in the K-subspace

Using the K-dimensional samples $\tilde{\boldsymbol{x}}_k$, the problem has moved from the case $N \gg K$ to $N = K$, where we can expect to have an accurate estimation of $\text{Tr}(\boldsymbol{R}_K)$ to be plugged in the α oracles as well as to be rid of the condition on α for the convergence of the RTE, thus being compatible with all oracles presented in section 2. In the subspace, the RTE is:

$$\hat{\boldsymbol{R}}_K = (1-\alpha)\sum_{k=1}^{K} \frac{\tilde{\boldsymbol{x}}_k\,\tilde{\boldsymbol{x}}_k^H}{\tilde{\boldsymbol{x}}_k^H\,\hat{\boldsymbol{R}}_K^{-1}\,\tilde{\boldsymbol{x}}_k} + \alpha\,\boldsymbol{I}_K. \tag{16}$$

In the following, the estimated α value (oracle) is denoted $\hat{\alpha}_0$.



(a) (b)

Fig. 1: (a) Alpha values when K/N varies. Thick gray line represents the limit. (b) Distance $\mathcal{D}$ when K/N varies. Parameters are $N = 153$, $\rho = 0.9$, $\nu = 0.5$, results are averaged over 1000 Monte-Carlo trials.

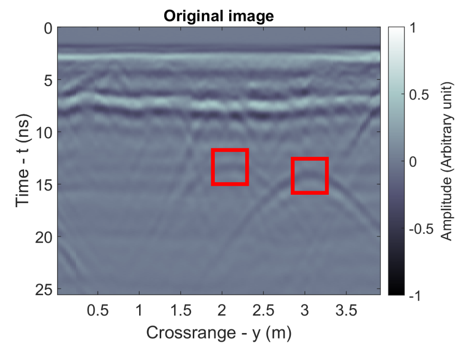


Fig. 2: Original GPR image. Objects locations are indicated by red squares.

3.3. Dimension expansion

After the computation of $\hat{\boldsymbol{R}}_K(\hat{\alpha}_0)$, the last operation is to shift back to the N-space and reconstruct the covariance matrix estimator $\hat{\boldsymbol{R}}$. As samples are lacking to estimate the orthogonal component $\hat{\boldsymbol{R}}_K^{\perp}$, we propose to replace it by a non informative solution $\alpha^{\perp}\boldsymbol{I}_{N-K}$. In this work, we propose to set $\alpha^{\perp} = \hat{\alpha}_0$ to preserve the the spectrum of the estimated matrix, since the lowest eigenvalue of $\hat{\boldsymbol{R}}_K$ is greater than $\hat{\alpha}_0$. $\hat{\boldsymbol{R}}$ is thus expressed as:

$$\hat{\boldsymbol{R}} = \left[\hat{\boldsymbol{U}}_K\,\hat{\boldsymbol{U}}_K^{\perp}\right] \begin{bmatrix} \hat{\boldsymbol{R}}_K(\hat{\alpha}_0) & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\alpha}_0\,\boldsymbol{I}_{N-K} \end{bmatrix} \left[\hat{\boldsymbol{U}}_K\,\hat{\boldsymbol{U}}_K^{\perp}\right]^H. \tag{17}$$

Note that other $\alpha^{\perp}$ values can be used in order to influence the properties of the estimator, e.g. regulate the PFA levels when used in an adaptive detection problem [18].

4. NUMERICAL SIMULATIONS

In this section we run simulations to illustrate the interest of the proposed subspace approach as compared to the RTE with oracles from section 2 in the undersampled case. To line up with the real data application of section 5, the simulations are run in the real case, and set up with $\{\boldsymbol{x}_k\}_{1,K}$ generated following a SIRV model, $\boldsymbol{x}_k = \sqrt{\tau_k}\,\boldsymbol{g}_k$, with $\boldsymbol{g}_k \sim \mathbb{R}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{R})$, $\boldsymbol{R}$ of Toeplitz form $[\boldsymbol{R}]_{ij} = \rho^{|i-j|}$, and $\tau \sim \Gamma(\nu, 1/\nu)$.

Fig. 1a shows the value of $\hat{\alpha}_0$ for both criteria in the undersampled case with and without subspace projection. The first remark is that the values for the CWH oracle in both cases are close to one another and under the convergence limit. This shows that though the subspace projection has no influence on the value of $\hat{\alpha}_0$, it now allows it to be used in algorithm (2). For the shape criterion, there is a gain in the estimated value, showing that the subspace projection allows to put less weight on the shrinkage and make more use of the data vectors.

Fig. 1b shows the difference between estimated matrices in terms of the average distance $\mathcal{D}(\boldsymbol{R}, \hat{\boldsymbol{R}})$ [1]:

$$\mathcal{D}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \|[N/\operatorname{Tr}(\boldsymbol{R}^{-1}\hat{\boldsymbol{R}})]\boldsymbol{R}^{-1}\hat{\boldsymbol{R}} - \boldsymbol{I}_N\|^2 . \qquad (18)$$

Results show a gain in using the subspace approach for low K/N ratios, with a decrease in performance when $K \to N$ as the subspace approach meets back with the traditional estimation. A similar study cannot be conducted for the CWH oracle since the $\hat{\alpha}_{CWH}$ values does not allow convergence when the subspace operation is not applied.

5. APPLICATION TO REAL DATA

Here, the proposed estimation scheme is applied to the adaptive detection problem in GPR images [19]. The problem consists in detecting a known signal $\boldsymbol{p} \in \mathbb{R}^N$ (response of a buried object) in an observation signal $\boldsymbol{x} \in \mathbb{R}^N$, while having a secondary set of response free observations $\{\boldsymbol{x}_k\}_{1,K}$ with $K \ll N$. Two hypotheses are formulated H_0: $\boldsymbol{x} = \boldsymbol{n}$ and H_1: $\boldsymbol{x} = a\,\boldsymbol{p} + \boldsymbol{n}$ and the problem boils down to the following detector:

$$\hat{\Lambda} = \max_{\epsilon' \in \mathbb{R}^+} \frac{|\boldsymbol{p}^T \hat{\boldsymbol{R}}^{-1}\boldsymbol{x}|^2}{(\boldsymbol{p}^T \hat{\boldsymbol{R}}^{-1}\boldsymbol{p})(\boldsymbol{x}^T \hat{\boldsymbol{R}}^{-1}\boldsymbol{x})} \underset{H_1}{\overset{H_0}{\lessgtr}} \eta , \qquad (19)$$

which is then applied to an image containing two objects (indicated by red squares) shown in fig. 2. Three estimators have been used: RTE with $\hat{\alpha}_{OTy}$ with and without subspace projection, RTE with $\hat{\alpha}_{CWH}$ with subspace operation. We use as a fourth estimator the RTE with the parameter $\hat{\alpha}_{PD}$ maximising the probability of detection through use of RMT [9].

Results on fig. 3 show an improvement in detection for the shape criterion, lowering the noise response level in the lower part of the image. The estimator using values from $\hat{\alpha}_{CWH}$ returns a further reduced noise response. It also yields a result similar to the maximum PD parameter $\hat{\alpha}_{PD}$ while being much faster to compute. To further compare the methods, a threshold at 20% of the maximum of each image is applied to compute the corresponding PFA levels. Results are $PFA_{OTy} = 1.15\,10^{-2}$, $PFA_{OTySub} = 2.1\,10^{-3}$, $PFA_{CWH} = 1.4\,10^{-3}$, $PFA_{PD} = 1.2\,10^{-3}$.

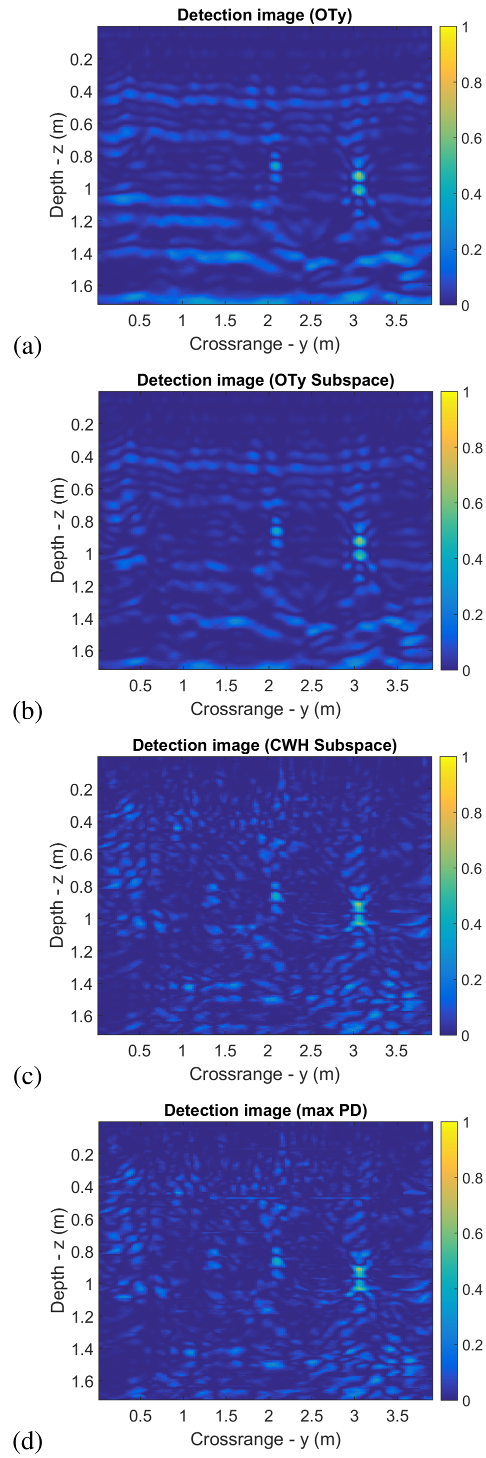


Fig. 3: Results for (a) α_{OTy} without subspace operation (range $\alpha \in [0.998; 1]$) (b) α_{OTy} with subspace operation (range $\alpha \in [0.8; 0.95]$), (c) α_{CWH} with subspace operation (range $\alpha \in [0.1; 0.4]$), and (d) maximum PD $\hat{\alpha}_{PD}$ value (range $\alpha \in [0.4; 0.9]$). Parameters are $N = 153$, $K = 40$

6. CONCLUSION

A new scheme for covariance matrix estimation in the undersampled case ($K \ll N$) has been proposed, relying on a K-subspace projection to estimate matrix core, and an expansion of the core in the original N-space. This approach negates convergence issues inherent to RTEs in this configuration and allows for a better estimation of the shrinkage parameter $\hat{\alpha}$ as illustrated by numerical simulations. Application to the adaptive detection problem shows promising results.

7. REFERENCES

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

[2] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, 2011.

[3] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, 2012.

[4] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5640–5651, Nov 2014.

[5] E. Ollila and D. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[6] Y. Sun, P. Babu, and D. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.

[7] R. Couillet and M. McKay, "Second order statistics of robust estimators of scatter. application to GLRT detection for elliptical signals," *Elsevier Journal of Multivariate Analysis*, vol. 131, pp. 99–120, 2014.

[8] L. Yang, R. Couillet, and M. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6684–6697, 2015.

[9] A. Kammoun, R. Couillet, F. Pascal, and i. M. Alouin, "Optimal design of the adaptive normalized matched filter detector," *(submitted to) IEEE Transactions on Information Theory, 2016, arXiv Preprint 1504.01252*.

[10] A. Kammoun, R. Couillet, F. Pascal, and M. Alouini, "Convergence and fluctuations of regularized tyler estimators," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1048–1060, 2016.

[11] R. Couillet, A. Kammoun, and F. Pascal, "Second order statistics of robust estimators of scatter. application to GLRT detection for elliptical signals," *Elsevier Journal of Multivariate Analysis*, vol. 143, pp. 249–274, 2016.

[12] N. Auguin, D. Morales-Jimenez, M. McKay, and R. Couillet, "Robust shrinkage m-estimators of large covariance matrices," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–4.

[13] T. Zhang and A. Wiesel, "Automatic diagonal loading for tyler's robust covariance estimator," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–5.

[14] Y. I. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 1: The over-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5807–5818, Dec 2013.

[15] O. Besson and Y. I. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 2: The under-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5819–5829, Dec 2013.

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

[17] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 3, April 2007, pp. III–1105–III–1108.

[18] O. Besson and Y. Abramovich, "Adaptive detection in elliptically distributed noise and under-sampled scenario," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1531–1535, Dec 2014.

[19] Q. Hoarau, G. Ginolhac, A. M. Atto, and J.-M. Nicolas, "Robust adaptive detection of buried pipes using GPR," *Signal Processing*, pp. –, 2016, available online.