TAPERED-SHRINKAGE COVARIANCE MATRIX ESTIMATION AND ITS APPLICATION TO HYPERSPECTRAL TARGET DETECTION

Yousef Zaitoun, Ahmad W. Bitar and Alexandre Baussard

LIST3N, Université de Technologie de Troyes, 12 rue Marie Curie, 10004 Troyes, France

ABSTRACT

Classical statistical hyperspectral target detectors rely on accurate covariance matrix estimation, particularly in high-dimensional settings. Tapering-based covariance estimators have improved such estimation in many applications. In this paper, we investigate the impact of different tapering weights within the TApered or BAnded Shrinkage COvariance matrix (TABASCO) framework. We demonstrate that both the exponential and wendland non-linear tapering weights can significantly improve the target detection performance in both Monte-Carlo simulations and hyperspectral data.

***Index Terms*—** Hyperspectral target detection, covariance estimation, TABASCO, tapering weights.

1. INTRODUCTION

In hyperspectral imagery, pixels are represented by vectors whose entries correspond to spectral bands, and images are represented by 3-D hypercubes. One of the most important applications of hyperspectral imagery is target detection, which can be viewed as a binary classification problem where pixels are labeled as target or background based on their spectral characteristics [1]. Its rich spectral information allows for more accurate material identification, and the main challenges lie in data variation modeling due to material spectral variability, atmospheric effects, and sensor noise.

Conventional target detection methods generally rely on the Gaussian assumption to model the background. Several statistical hypothesis testing techniques have been proposed in the literature to detect various targets of interest in hyperspectral imagery [2]. However, the detection performance of these classical detectors [1, 2] mainly depends on the true unknown covariance matrix (of the background surrounding the test pixel), whose entries have to be carefully estimated, especially in large dimensions.

As the capacity to measure and collect data increases, high dimensional signals and systems have become more and more prevalent. Hyperspectral imagery analysis is an example of areas in which the dimensionality of signals is growing explosively. This growth presents new challenges in the modeling and analysis of high dimensional data. In hyperspectral imagery, the dimensionality of signals often grows much faster than the number of available secondary data, and hence, it becomes impractical to use traditional covariance estimators where the target detection performance of the classical target detectors can deteriorate significantly. Usually, the researchers assume that compounding the high dimensionality problem can be alleviated via tapering and shrinkage techniques.

This paper exploits the use of the TApered or BAnded Shrinkage COvariance matrix (TABASCO) estimator [3] for hyperspectral target detection in the case when the number of spectral bands is higher than the number of secondary data used to estimate the true unknown covariance matrix. This estimator mainly shrinks the tapered sample covariance matrix towards a scaled identity matrix after choosing a specific linear tapering weight. Our main contribution in this paper lies in replacing the conventional linear tapering weights by non-linear tapering ones designed to better capture the correlation structure of hyperspectral imagery.

2. BACKGROUND AND SYSTEM OVERVIEW

Consider each pixel in the hyperspectral image, $\mathbf{x} \in \mathbb{R}^p$, as a p-dimensional vector that follows a Gaussian distribution with unknown mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$, where p designates the total number of spectral bands. Given n secondary data $\{\mathbf{x}_i\}_{i=1}^n$ used for local covariance estimation, the traditional (unbiased) estimator of $\boldsymbol{\Sigma}$ is the sample covariance matrix defined as

$$\hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \ , \qquad (1)$$

where $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$. When $p > n$, the estimator $\hat{\boldsymbol{\Sigma}}_{SCM}$ is ill-conditioned and not invertible. However, the problem of estimating high-dimensional covariance matrices has been extensively studied. Realizing the challenges brought by the high covariance dimensionality, researchers have proposed various regularization techniques in many applications, such as shrinkage and tapering.

Chen et al. [4] further developed the concepts of tapering and shrinkage by enhancing the estimator through their proposal of the shrinkage-to-tapering oracle approximation

(STOA), which is based on the Gaussian assumption. Subsequently, Zhou et al. [5] studied the STOA under both Gaussian and non-Gaussian data. The limitation of this study is that in high-dimensional setting, the covariance matrix estimations are not necessarily positive semi-definite. This restriction prevents any guaranty for a regularized positive definite estimation, which is essential for many statistical hyperspectral imagery and detection applications.

The tapering of $\hat{\boldsymbol{\Sigma}}_{SCM}$ is defined as $\boldsymbol{W} \circ \hat{\boldsymbol{\Sigma}}_{SCM}$, where $\circ$ represents the Hadamard element-wise product [6] and $\boldsymbol{W} = [w_{ij}]_{p\times p}$ is the tapering weight matrix, where $w_{ij} \in [0,\, 1]$. The role of $\mathbf{W}$ is to control how strongly to keep each covariance entry. By assuming that $w_{ij} = g(x)$, the function g should be continuous, symmetric, positive definite, satisfy $g(0) = 1$, and decreasing. Such a positive-definite linear tapering weight which considers that entries different than the diagonal are less important [7, 8], can be defined as

$$w_{ij}^{(k)} = \begin{cases} 1, & |i-j| < k \\ 0, & |i-j| \geq k\,, \end{cases} \tag{2}$$

where $k \in [1,\, p]$ is the bandwidth parameter which controls the sparsity level, and $|i-j|$ is the spectral bands difference. Another linear tapering weight that keeps the entries near the diagonal fully, then decreases linearly, and finally sets far entries by zero [9], can be defined as

$$w_{ij}^{(k)} = \begin{cases} 1, & |i-j| \leq \frac{k}{2} \\ 2 - \dfrac{2|i-j|}{k}, & \frac{k}{2} < |i-j| < k\,. \\ 0, & |i-j| \geq k \end{cases} \tag{3}$$

The linear tapering weight avoids an abrupt cutoff of jumping directly from 1 to 0; this usually gives a smoother covariance estimation. In addition, the choice of k clearly affects the performance of the tapering matrix, where the optimal choice of k depends on measuring the norm of the estimation error.

Ollila et al., in [3], proposed the TABASCO estimator, which combines tapering weight that imposes the structure of the covariance matrix and shrinkage toward a scaled identity matrix that stabilizes the estimator when the number of secondary data is small. This work motivates the estimation of the true unknown covariance matrix $\boldsymbol{\Sigma}$, especially under the elliptical distribution, where Ollila proved in the experiments the advantage of TABASCO compared to the state-of-the-art as STOA approach. Ollila focused on the conventional linear tapering weights (that is, in (2) and (3)) in the simulation studies, while the Kronecker sinc tapering weights were used for real space-time adaptive processing (STAP) experiments. The main limitation of [3] of the Kronecker sinc tapering weights that we found out is the computational burden for hyperspectral imagery. In addition, the results of the linear tapering weights were implemented only in sparse covariance settings, which may fail in other scenarios.

In this paper, we propose to replace the conventional linear tapering weights in (2) and (3) by two non-linear ones that we expect to improve the target detection performance in hyperspectral imagery. Our two proposed tapering weights, in section 3, are evaluated for hyperspectral target detection using the normalized matched filter. More precisely, we first perform a thorough evaluation of the two proposed tapering weights on some Monte-Carlo simulations for three true covariance matrix models of different sparsity levels in subsection 4.1. From our simulations, the target detection results show that in truly sparse models, our two proposed tapering weights achieve almost the same detection performance as the conventional linear ones. However, when the true covariance model is not sparse or dense, we find that empirically our two proposed weights significantly improve the target detection performance compared to the conventional linear weights. Next, in subsection 4.2, our two proposed weights are evaluated on synthetic data, the results of which demonstrate their effectiveness in detecting targets with low target fill-fraction. Finally, section 5 gives concluding remarks and some directions for future work.

3. MAIN CONTRIBUTION

Before describing the proposed two non-linear tapering weights, we recall the TABASCO estimator that jointly leverages shrinkage to identity and tapering as

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{W}} = \beta\big(\boldsymbol{W}(k) \circ \hat{\boldsymbol{\Sigma}}_{SCM}\big) + (1-\beta)\frac{Tr\left(\hat{\boldsymbol{\Sigma}}_{SCM}\right)}{p}\boldsymbol{I}\,,$$

where $Tr\,(.)$ stands for the trace, and $\beta \in [0, 1]$ is the shrinkage coefficient that controls the balance between the structured covariance term and the identity matrix. However, if β is close to 1, the algorithm trusts the tapered covariance estimate more strongly. If β is close to 0, the algorithm prefers a stronger shrinkage toward the identity matrix. Both the parameters β and k are computed by jointly minimizing the mean square error $\mathbb{E}\left(\left\|\hat{\boldsymbol{\Sigma}}_{\mathbf{W}} - \boldsymbol{\Sigma}\right\|_F^2\right)$ under the assumption of elliptically symmetric distribution [10].

All parameters used in TABASCO and the mathematical derivations of both β and k, are detailed in [3]. In this work, we adopt the TABASCO framework and focus on the design of alternative tapering weight matrices $\boldsymbol{W}$ to improve the covariance matrix estimation for hyperspectral target detection. More precisely, we propose to consider two non-linear tapering weights such as the wendland and exponential.

3.1. The wendland tapering weight

The wendland weight uses a polynomial nonlinear form to reduce the complexity of large covariance matrices. The wendland tapering weights are widely used in spatial statistics and

Gaussian process modeling [11]. They have already been used for covariance tapering estimation, but to our knowledge, never in the context of hyperspectral target detection. We consider the following wendland tapering weight [12]:

$$w_{ij}^{(k)} = \begin{cases} \left(1 - \frac{|i-j|}{k}\right)^4 \left(1 + 4\frac{|i-j|}{k}\right), & |i-j| \leq k \\ 0, & |i-j| > k\,. \end{cases} \quad (4)$$

3.2. The exponential tapering weight

We aim to develop the exponential tapering weight, only mentioned in Bickel and Levina [7], which satisfies the weight-tapering properties. We consider the following weight:

$$w_{ij}^{(k)} = \exp\left(-\frac{|i-j|}{k}\right). \quad (5)$$

In this context, the tapering parameter k chosen by TABASCO implementation will control the decay speed where a small k gives fast decay, otherwise a slow decay. To our knowledge, the exponential weight has not been used before neither for tapering covariance estimation nor in the context of hyperspectral target detection.

3.3. Our main motivation

In hyperspectral imagery, adjacent spectral bands are highly correlated because they correspond to neighboring wavelengths measuring the same physical materials. This motivates our contribution to investigate a suitable covariance non-linear tapering weights to improve the target detection. The wendland tapering weight is mainly motivated by its ability to improve the covariance estimation by creating sparse covariance matrices, which greatly preserves local spatial dependencies while maintaining positive definiteness and reducing estimation variance. On the other hand, the exponential weight reflects the good transition in spectral responses that are continuous and smooth across wavelengths. Furthermore, the spectral correlations typically decay smoothly where nearby covariance coefficients receive a large weight, which means neighboring spectral bands are highly correlated. Unlike hard-band tapering, which removes distant covariance entries, exponential and wendland tapering progressively preserve long-range information. Accordingly, both tapering weights aim to improve covariance estimation, reduce the variance associated with high-dimensional covariance estimation, and achieve a better bias–variance tradeoff.

Figure 1 compares the wendland and exponential tapering weights with the two conventional linear tapering ones in (2) and (3). As we can observe, the weight in (2) introduces a discontinuous weighting function and additional bias if the true covariance matrix contains significant correlations beyond the selected band. However, the weight in (3) assumes that correlation will decrease linearly with distance and equally without considering the actual correlation structure. On the other hand, the wendland weight provides smooth nonlinear attenuation of covariance coefficients. In contrast, exponential weight eliminates the need for a threshold, which provides smoothness, monotonicity, and continuity. Neither the Wendland nor the exponential tapering weights have previously been considered in the TABASCO framework and especially for hyperspectral target detection. Therefore, this work compares the two considered non-linear tapering weights with the conventional linear tapering ones in (2) and (3) to evaluate the detection performance in hyperspectral imagery.

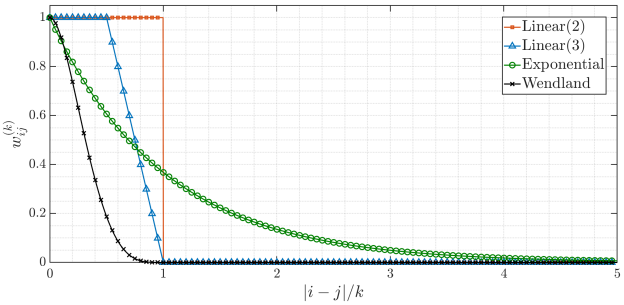


Fig. 1. Comparison of tapering weights.

4. HYPERSPECTRAL TARGET DETECTION

In this section, we perform both Monte-Carlo simulations as well as synthetic experiments to gauge the target detection performance of our proposed non-linear tapering weights in subsections 3.1 and 3.2. We mainly focus on the case when $p > n$. The target detection is formulated as a hypothesis testing problem to determine whether each test pixel contains the target of interest or not. To do this, we employ the Adaptive Normalized Matched Filter (ANMF) [13]:

$$\text{ANMF}(\mathbf{x}) = \frac{\left|\mathbf{t}^\top \hat{\mathbf{\Sigma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})\right|^2}{\left(\mathbf{t}^\top \hat{\mathbf{\Sigma}}^{-1} \mathbf{t}\right)\left((\mathbf{x} - \hat{\boldsymbol{\mu}})^\top \hat{\mathbf{\Sigma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta\,,$$

where $\mathbf{x} \in \mathbb{R}^p$ is the test pixel and $\mathbf{t} \in \mathbb{R}^p$ is the known target signature. The threshold $\delta > 0$ is selected according to the desired probability of false alarm (P_{fa}): the null hypothesis $\mathcal{H}_0$ (target absent) is accepted for detectors that does not exceed δ. Otherwise, the alternative hypothesis $\mathcal{H}_1$ (target present) is accepted. We denote the covariance matrices estimated using our proposed non-linear wendland and exponential tapering weights in (4) and (5) by $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$, respectively. The following covariance matrices will be used in the comparison:

- $\hat{\mathbf{\Sigma}}_{\text{SCM}}$: the sample covariance matrix in (1);
- $\hat{\mathbf{\Sigma}}_{\text{Linear(2)}}$ and $\hat{\mathbf{\Sigma}}_{\text{Linear(3)}}$: the TABASCO estimator using the tapering weight in (2) and (3), respectively.

Note that in the Monte-Carlo simulations only, all the covariance matrices above will be further compared with the true

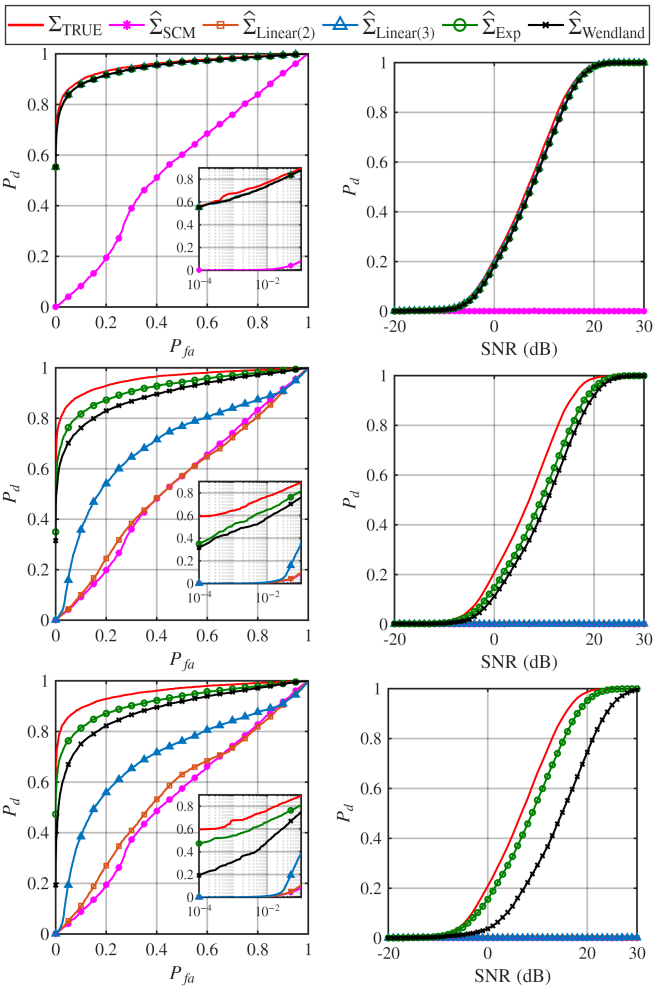


Fig. 2. Receiver operating characteristic (ROC) curves, in terms of (P_d, P_{fa}) and (P_d, SNR), of the different covariance estimators under different covariance structures. From top row to bottom row: AR(1) covariance model, triangular covariance model with $r = p/2$, triangular covariance model with $r = p$.

known covariance matrix, denoted as $\mathbf{\Sigma}_{TRUE}$, which should obviously achieve the best target detection performance.

4.1. Monte-Carlo simulations

In this section, we perform 10^4 Monte-Carlo trials using all TABASCO estimation settings with the elliptically distributed background generated under the sphericity of Ell1-estimator (for more details, see subsection V-A in [3]). We fix $n = 30$, $p = 100$, and the secondary data are generated from a p-variate K-distribution with shape parameter equal to 0.5. The experiments are conducted on the following true covariance matrices:

- The Autoregressive model order 1, AR(1):

$$\mathbf{\Sigma}_{\text{TRUE}} = [\sigma_{ij}]_{p\times p}, \quad \sigma_{ij} = c^{|i-j|}, \quad c = 0.3\,.$$

- The triangular matrix with $r = p/2$ and $r = p$:

$$\mathbf{\Sigma}_{\text{TRUE}} = [\sigma_{ij}]_{p\times p}, \quad \sigma_{ij} = \left(1 - \left(\frac{|i-j|}{r}\right)\right)_+ .$$

The AR(1) is approximately sparse. The triangular matrix with $r = p/2$ is much less sparse than AR(1), whereas the one with $r = p$ is very dense.

Figure 2 confirms that both $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$ significantly improves the target detection performance especially when the true covariance matrix is not sparse. However, $\hat{\mathbf{\Sigma}}_{\text{Exp}}$ achieves the highest probability of detection (P_d). Because of the significant sparsity displayed in AR(1), $\hat{\mathbf{\Sigma}}_{\text{Linear(2)}}$ and $\hat{\mathbf{\Sigma}}_{\text{Linear(3)}}$ perform similarly to $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$. On the other hand, the triangular model with both $r = p/2$ and $r = p$, demonstrates that $\hat{\mathbf{\Sigma}}_{\text{Exp}}$ outperforms the other estimators. Since $p > n$, $\hat{\mathbf{\Sigma}}_{SCM}$ performs quite badly: its flat P_d–SNR curves match its low P_d at fixed P_{fa} in the ROC curves. More importantly, $\hat{\mathbf{\Sigma}}_{\text{Linear(2)}}$ and $\hat{\mathbf{\Sigma}}_{\text{Linear(3)}}$ perform poorly with the triangular covariance for both $r = p/2$ and $r = p$.

4.2. Application on synthetic data

In this section, we perform synthetic experiments on a $101 \times 101 \times 186$ zone acquired from the AVIRIS Cuprite dataset [14]. We incorporate in this zone 7 target blocks (each of size 3×3, a total of 63 target pixels), placed in long convoy formation (see the top left of figure 3), and all formed by the same synthetic sulfate mineral type known as "Jarosite" selected from the United States Geological Survey (USGS) spectral library. The seven target blocks are embedded to replace a fraction from the background at the same spatial locations[1]. This fraction is fixed to 0.3. The target signature $\mathbf{t}$ in the ANMF detector consists of the mean of six Jarosite mineral samples taken from the USGS spectral library. For each test pixel in the hyperspectral image, the secondary data are constructed using a 9×9 outer window region and a 5×5 inner window region. Hence, we obtain a total of $n = 56$ secondary data and $p = 186$ spectral bands. From the top right of figure 3, we can observe that $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ achieves a slightly better target detection than $\hat{\mathbf{\Sigma}}_{\text{Exp}}$, where both significantly outperform $\hat{\mathbf{\Sigma}}_{\text{Linear(2)}}$ and $\hat{\mathbf{\Sigma}}_{\text{Linear(3)}}$ especially at low P_{fa} values. More precisely, $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$ achieve the best target detection performance at $P_{fa} = 10^{-4}$ with $P_d = 0.905$ and $P_d = 0.873$, respectively. Furthermore, both $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$ and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$ are more effective and consistent in detecting the embedded seven target blocks especially for very low P_{fa}, as shown in the visual detection results of figure 3.

5. CONCLUSION AND FUTURE WORK

In this paper, we compare different tapering weights among the TABASCO covariance estimation framework for hyper-

[1]Note that both the hyperspectral image and the Jarosite target samples are normalized to values between 0 and 1.

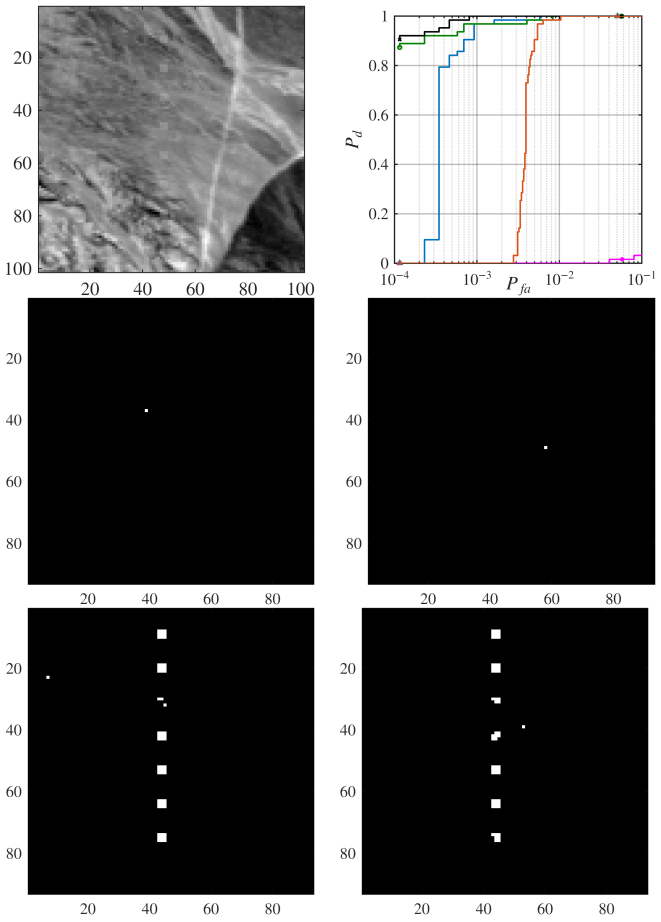


Fig. 3. From top left to bottom right: the original $101 \times 101 \times 186$ pixel (mean power in dB) with the 7 embedded targets blocks at a target fill-fraction equal to 0.3; ROC curves (same legend as in Fig.2); and thresholded detection maps using $\hat{\mathbf{\Sigma}}_{\text{Linear(2)}}$, $\hat{\mathbf{\Sigma}}_{\text{Linear(3)}}$, $\hat{\mathbf{\Sigma}}_{\text{Wendland}}$, and $\hat{\mathbf{\Sigma}}_{\text{Exp}}$, respectively. The detector output is of size 93×93 pixels.

spectral target detection when $p > n$. Both the Monte-Carlo simulations and synthetic experiments proved significant improvement in the detection performance for the wendland and exponential tapering weight when compared to conventional linear weights, particularly for low probability of false alarms. Regarding future enhancements, we aim to evaluate the different tapering weights on more challenging datasets and especially for targets with lower target fill-fractions.

6. REFERENCES

[1] Dimitris Manolakis, David Marden, Gary A Shaw, et al., "Hyperspectral image processing for automatic target detection applications," *Lincoln laboratory journal*, vol. 14, no. 1, pp. 79–116, 2003.

[2] Dimitris Manolakis, Eric Truslow, Michael Pieper, Thomas Cooley, and Michael Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 24–33, 2013.

[3] E. Ollila and A. Breloy, "Regularized tapered sample covariance matrix," *IEEE Trans. Signal Process.*, vol. 70, pp. 2306–2320, 2022.

[4] X. Chen, Z. J. Wang, and M. J. McKeown, "Shrinkage-to-tapering estimation of large covariance matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5640–5656, 2012.

[5] J. Li, J. Zhou, and B. Zhang, "Estimation of large covariance matrices by shrinking to structured target in normal and non-normal distributions," *IEEE Access*, vol. 6, pp. 2158–2169, 2017.

[6] R. Furrer and T. Bengtsson, "Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants," *Journal of Multivariate Analysis*, vol. 98, no. 2, pp. 227–255, 2007.

[7] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[8] P. J. Bickel and E. Levina, "Covariance regularization by thresholding," *The Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.

[9] T. T. Cai, C. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *The Annals of Statistics*, vol. 38, no. 4, pp. 2118–2144, 2010.

[10] K. Fang, S. Kotz, and K. W Ng, *Symmetric multivariate and related distributions*, Chapman and Hall/CRC, 2018.

[11] C. G Kaufman, M. J Schervish, and D. W Nychka, "Covariance tapering for likelihood-based estimation in large spatial data sets," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1545–1555, 2008.

[12] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Advances in computational Mathematics*, vol. 4, no. 1, pp. 389–396, 1995.

[13] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A cfar adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[14] A. W. Bitar., L-F. Cheong., and J-P. Ovalez., "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing.*, vol. 57, no. 8, pp. 5239–5251, 2019.