Online change detection in SAR time-series with Kronecker product structured scaled Gaussian models

Ammar Mian, Guillaume Ginolhac, Florent Bouchard, Arnaud Breloy

Online Change Detection in SAR Time-Series with Kronecker Product Structured Scaled Gaussian Models

Ammar Mian, Guillaume Ginolhac[a], Florent Bouchard[c], Arnaud Breloy[1]

[a]*LISTIC, University Savoie Mont-Blanc, 5 Chemin de bellevue, Annecy, 74940, France*
[b]*L2S, CNRS, CentraleSupelec, University Paris-Saclay, 3 Rue Joliot-Curie, Gif-sur-Yvette, 91190, France*
[c]*LEME, University Paris Nanterre, 50 Rue de Sevres, Ville-d'Avray, 92410, France*

Abstract

We develop the information geometry of scaled Gaussian distributions for which the covariance matrix exhibits a Kronecker product structure. This model and its geometry are then used to propose an online change detection (CD) algorithm for multivariate image times series (MITS). The proposed approach relies mainly on the online estimation of the structured covariance matrix under the null hypothesis, which is performed through a recursive (natural) Riemannian gradient descent. This approach exhibits a practical interest compared to the corresponding offline version, as its computational cost remains constant for each new image added in the time series. Simulations show that the proposed recursive estimators reach the Intrinsic Cramér-Rao bound. The interest of the proposed online CD approach is demonstrated on both simulated and real data.

Keywords: Online covariance matrix estimation, Riemannian Geometry, Scaled Gaussian, Change Detection

1. Introduction

Synthetic Aperture Radar (SAR) Multivariate Image Times Series (MITS) have lately been made widely available thanks to various Earth monitoring missions such as Sentinel-1, TerraSAR-X, or UAVSAR. This modality offers significant advantages over the optical and multispectral ones, notably when the area of interest is observed at night, or covered by clouds. As for optical images, the pixel of SAR images can be multivariate when exploiting

the polarimetric diversity, or the spectro-angular properties of the scatterers through a wavelet decomposition [1]. Hence, the analysis of multivariate SAR-MITS has become an active topic of research over the past years, with numerous applications such as change detection and crop classification.

In this paper, we focus on change detection applications for SAR-MITS [2]. Notably, when the local patches of the images are used instead of single pixels, the detection algorithms are mainly based on the analysis of the statistical properties of the patch over the time. In this framework, a popular approach is based on testing the change of the covariance matrix along the time series [3]. An appropriate matrix distance [4, 5, 6] is then used to take a decision. For time series with more than two images, a prominent class of methods builds this distance upon the Generalized Likelihood Ratio Test (GLRT), usually constructed under the assumption that the multivariate pixels follow a Gaussian distribution [7, 8]. In practice, the radar returns makes the data distribution heavy-tailed [9, 10]. A popular family to model non Gaussian radar data is the Complex Elliptically Symmetric (CES) distributions framework [11], that encompasses a significant number of well-known distribution (e.g., Weibull, Student-t, and K-distribution). Scaled Gaussian (SG) distributions are a major sub-family of CES, that model each sample as Gaussian conditionally to a scale factor called the texture. The choice of the texture distribution allows then for accurately modeling the tails of the distribution. Conversely, assuming that the texture is unknown and deterministic for each sample offers a more robust model, as it is distribution-free [12, 13]. This particular model has been used in [14] to extend the classical GLRT of [7, 8], which improved the detection performance when using high resolution SAR images.

The performance of the aforementioned change detection methods is generally improved when the input data reflects the spectro-angular diversity of the scatterers. Such diversity can be accounted for through an appropriate wavelet transformation [1]. This pre-processing incidentally increases the pixel dimension p compared to the native polarimetric SAR pixels (for which $p = 3$). As a result, the spatial resolution of covariance-based change detectors is reduced, as the tested local patches should include a number of pixel n that scales with p. This issue can be mitigated by assuming an additional structure for the covariance matrix and derive the corresponding change detection algorithm. This approach was successfully leveraged by considering low-rank structures of the covariance matrix in [15, 16]. A limitation regarding these methods is that the rank of the covariance matrix is

an extra parameter that needs to be estimated. Moreover, this rank often changes with respect to the studied pixel and the images which leads to complicated strategies for the change detection process. This motivates the use of alternate low-dimensional covariance matrix structures.

In this work, we remark that the transformation proposed in [1] combined with polarization channels naturally leads to have an inherent Kronecker product structure. This structure is interesting since it does not involve any nuisance parameter: the dimensions of the sub-matrices are fixed by the number of polarization channels. It also greatly reduces the number of parameters to be estimated, which allows for preserving a thin spatial resolution. The Kronecker product structure has been extensively leveraged in the radar community and for MIMO systems. For example, several algorithms have been developed for the estimation of Kronecker product structured covariance matrices when the distribution is assumed to be Gaussian [17, 18]. Their extensions to Scaled Gaussian (SG) model, have been developed in [19, 20]. To the best of our knowledge, this structure was not considered in change detection for SAR-MITS. Hence, building upon the aforementioned works, the first main contribution of this paper is thus to derive the GLRT for the change detection problem when the data follows a SG distribution with a Kronecker structure for the covariance matrix.

We then address important issues regarding the computational load of robust change detection methods. Indeed, the computational cost of robust change detection [14, 16] becomes prohibitive when the number of images T increases, which limits their practical use for large SAR-MITS. The main computational bottleneck is the computation of a robust covariance matrix estimate, that needs to be re-evaluated under the null hypothesis for each new image. This issue can be mitigated by replacing the estimate by one computed online (i.e., in a one pass streaming process). In this scope, the Riemannian optimization framework [21] offers an interesting lead since the covariance matrix is a parameter that belongs to a smooth manifold: the manifold of Hermitian Positive Definite (HPD) matrices. Especially, a Riemannian gradient based online procedure proposed in [22] allows for obtaining a statistically efficient estimate in a computationally efficient manner (as the gradient steps do not require any grid search). This method has already been successfully leveraged for scaled Gaussian models in [23] and t-distributed data in [24]. The second main contribution of this paper is to leverage this framework for reducing the computational load of the proposed change detection method. This requires to study the information geometry of

scaled Gaussian models with Kronecker product structured covariance matrices. As a side result of this analysis, we also obtain the intrinsic Cramér-Rao bound [25, 26] for the corresponding estimation problem. This bound is then used to validate the proposed online estimation method on simulated data.

Finally, we demonstrate the interest of the proposed approach on a real dataset provided by UAVSAR (courtesy of NASA/JPL-Caltech). The SAR-MITS is referenced as SDelta_28518_02, Segment 1, which shows the evolution of a river delta in the USA with cycle of droughts and flood.

The rest of the article is organized as follows: Section 2 presents the data model and the GLRT adapted to a Kronecker product structured covariance matrix. Section 4 studies the information geometry of this model, and presents the derivation of the corresponding online estimation algorithm. Section 5 illustrates the performance of the proposed method on simulated and real data. Notations: italic indicates a scalar quantity, lower case boldface indicates a vector quantity, and upper case boldface a matrix. The transpose conjugate operator is ^{H} and the conjugate one is *. $tr(\cdot)$ and $|\cdot|$ are respectively the trace and the determinant operators. $\mathcal{H}_p^{++}$ is the manifold of HPD matrices of size $p \times p$. For $\mathbf{x} \in \mathbb{C}^p$, the notation $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ stands for a complex-valued random Gaussian vector of mean $\boldsymbol{\nu} \in \mathbb{C}^p$ and covariance matrix $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$.

2. A framework for robust covariance matrix change detection

2.1. Setup

Within a SAR-MITS, we consider detecting a change within a local (spatial) patch of pixel. Given a set of T co-registered images, this means processing T sets of n pixels of dimension p, denoted $\{\{\boldsymbol{x}_i^{(t)}\}_{i\in[\![1,n]\!]}\}_{t\in[\![1,T]\!]}$. The corresponding setup is displayed in Figure 1. A change detection procedure can then be derived by: *i*) expressing a relevant statistical model and corresponding likelihood function for the dataset; *ii*) deriving a change detection test (such as the GLRT) to assess whether the parameters of the statistical model change over time, or not. In the context of SAR-MITS, such procedures usually test for a change within the covariance matrix of the patch, as it is a meaningful feature for this type of data [3, 4, 5, 6, 7, 8, 14].

2.2. Robust statistical model for the pixel patches

For a given time t the pixel patch consists of n i.i.d. samples $\{\boldsymbol{x}_i^{(t)}\}_{i\in[\![1,n]\!]}$ where each sample $\boldsymbol{x}_i^{(t)} \in \mathbb{C}^p$. Most work modeled this data as follow-

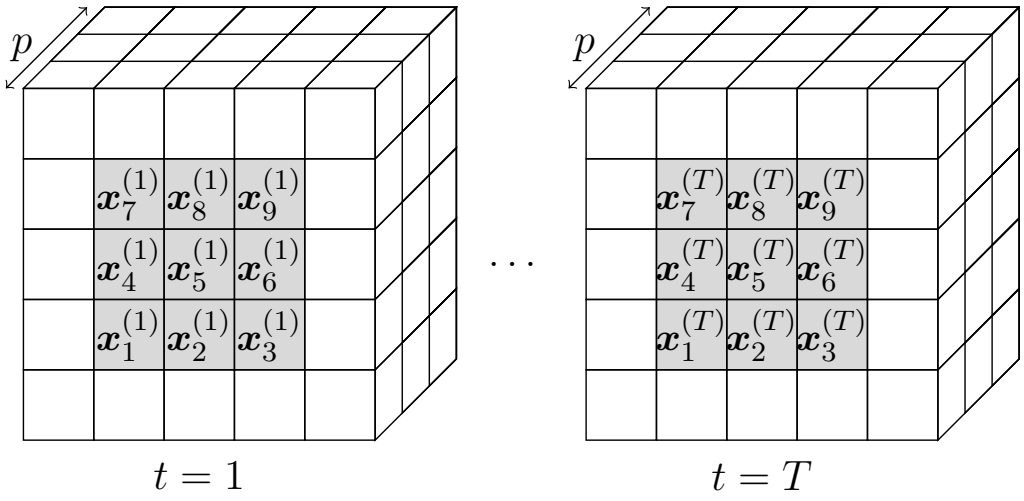


Figure 1: Illustration data notation for change detection test in a MITS.

ing a complex Gaussian distribution [7, 8], i.e., that pixels are drawn from $\boldsymbol{x}_i^{(t)} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}^{(t)})\ \forall i$, with $\boldsymbol{\Sigma}^{(t)} \in \mathcal{H}_p^{++}$. However, it was shown in [14] that the complex SG distribution offers a better fit to high resolution SAR images, and allows for improving the performance of change detection. In this case each sample $\boldsymbol{x}_i^{(t)}$ is modeled as complex Gaussian conditionally to an unknown scale, referred to as the texture. The choice of the texture distribution allows then for modeling many well known heavy-tailed distributions [11]. Following from the robust estimation framework [12, 13], a distribution-free detection process will be obtained by assuming that the texture is unknown and deterministic for each sample. Hence, we model samples as $\boldsymbol{x}_i^{(t)} \sim \mathcal{CN}(\mathbf{0}, \tau_i^{(t)}\boldsymbol{\Sigma}^{(t)})$ with $\tau_i^{(t)} \in \mathbb{R}_*^+, \forall\ i \in [\![1, n]\!]$. The log-likelihood of the pixel patch at time t is then, up to a constant,

$$\mathcal{L}(\{\boldsymbol{x}_i^{(t)}\}_{i\in[\![1,n]\!]}, \theta^{(t)}) = \sum_{i=1}^{n} \log|\tau_i^{(t)}\boldsymbol{\Sigma}^{(t)}| + \frac{1}{\tau_i^{(t)}}\, \boldsymbol{x}_i^{(t)\,H}\, \boldsymbol{\Sigma}^{(t)\,-1}\, \boldsymbol{x}_i \qquad (1)$$

where $\theta^{(t)} = \{\boldsymbol{\Sigma}^{(t)}, \boldsymbol{\tau}^{(t)}\}$ is the set of unknown statistical parameters, in which $\boldsymbol{\tau}^{(t)} = [\tau_1^{(t)},\ \cdots,\ \tau_n^{(t)}]$ denotes the vector of textures. Though this will not be exploited in this work, we also notice that the Gaussian model can be recovered as a sub-case by simply setting $\tau_i^{(t)} = 1, \forall\ i \in [\![1, n]\!]$.

2.3. Change detection with the GLRT

Covariance based change detection operates by assuming that a change in the image is translated by a change in the unknown parameters $\theta^{(t)}$ over the time index t. From the SAR-MITS $\{\{\boldsymbol{x}_i^{(t)}\}_{i\in[\![1,n]\!]}\}_{t\in[\![1,T]\!]}$, the change detection problem can thus be written as the following hypothesis test:

$$\begin{cases} \mathrm{H}_0 : \theta^{(1)} = \theta^{(2)} = ... = \theta^{(T)} \triangleq \theta^{(0)} \\ \mathrm{H}_1 : \exists (t,t') \in [\![1,T]\!]^2,\ \theta^{(t)} \neq \theta^{(t')} \end{cases} \tag{2}$$

For the SG model of Section 2.2, the GLRT corresponding to this hypothesis test has been derived in [14] and is expressed as:

$$\hat{\Lambda}_{SG} = \frac{\left|\hat{\boldsymbol{\Sigma}}^{(0)}\right|^{Tn}}{\prod_{t=1}^{T}\left|\hat{\boldsymbol{\Sigma}}^{(t)}\right|^{n}} \prod_{i=1}^{n} \frac{\left(\hat{\tau}_i^{(0)}\right)^{Tp}}{\prod_{t=1}^{T}\left(\hat{\tau}_i^{(t)}\right)^{p}} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda, \tag{3}$$

where $\hat{\theta}^{(0)} = \{\hat{\boldsymbol{\Sigma}}^{(0)}, \hat{\boldsymbol{\tau}}^{(0)}\}$ stands for the maximum likelihood estimator (MLE) of the parameters under H_0, and $\hat{\theta}^{(t)} = \{\hat{\boldsymbol{\Sigma}}^{(t)}, \hat{\boldsymbol{\tau}}^{(t)}\},\ \forall t \in [\![1,T]\!]$ for the MLE of the parameters under H_1. These MLEs are expressed respectively as

$$\hat{\boldsymbol{\Sigma}}^{(0)} = \frac{p}{n}\sum_{i=1}^{n} \frac{\sum_{t=1}^{T} \boldsymbol{x}_i^{(t)}\, \boldsymbol{x}_i^{(t)\,H}}{\sum_{t=1}^{T} \boldsymbol{x}_i^{(t)\,H}\, \hat{\boldsymbol{\Sigma}}^{(0)\,-1}\, \boldsymbol{x}_i^{(t)}} \quad \text{and} \quad \hat{\tau}_i^{(0)} = \sum_{t=1}^{T} \frac{\boldsymbol{x}_i^{(t)\,H}\, \hat{\boldsymbol{\Sigma}}^{(0)\,-1}\, \boldsymbol{x}_i^{(t)}}{Tp}, \tag{4}$$

and

$$\hat{\boldsymbol{\Sigma}}^{(t)} = \frac{p}{n}\sum_{i=1}^{n} \frac{\boldsymbol{x}_i^{(t)}\, \boldsymbol{x}_i^{(t)\,H}}{\boldsymbol{x}_i^{(t)\,H}\, \hat{\boldsymbol{\Sigma}}^{(t)\,-1}\, \boldsymbol{x}_i^{(t)}} \quad \text{and} \quad \hat{\tau}_i^{(t)} = \frac{\boldsymbol{x}_i^{(t)\,H}\, \hat{\boldsymbol{\Sigma}}^{(t)\,-1}\, \boldsymbol{x}_i^{(t)}}{p}, \tag{5}$$

with the latter corresponds to Tyler's M-estimator of the scatter matrix [12, 13].

2.4. Some practical limitations

Data dimension p (pixel depth): For SAR-MITS, the diversity in the data arises from polarization channels (HH, VV and HV/VH), and the spectro-angular diversity of the scatterers which can be obtained by using an appropriate wavelet transform [1]. The latter transformation tends to improve the

detection performance of the covariance based change detectors (such as $\hat{\Lambda}_{SG}$ in (3)), however it increases the pixel depth p. It is consequently detrimental to the spatial resolution of the detectors, as the number of pixel in the patch n needs to scale with p to accurately estimate the local covariance matrix. The assumption of i.i.d. data can also be violated when processing large areas, which motivates the development of methods that reduce the number of needed samples. This reduction can be achieved by using regularization [27], or low-rank structured covariance matrices [16]. Nevertheless these methods involve regularization parameters that need proper tuning, which increases the complexity of the detection process. This issues will be addressed in Section 3, where we leverage a Kronecker product structure whose hyperparameters are inherently set by the wavelet transform [1].

Data dimension T (length of the time series): The detector $\hat{\Lambda}_{SG}$ in (3) involves solution of fixed point equations (5) and (4). Especially, the fixed point in (4) depends on the whole dataset and cannot be computed recursively (i.e., in a streaming fashion). Whether a low-dimensional structure is used or not, this means that the computational load of the robust detectors increases heavily for each new added image in the stack, which can be limiting. This issue is addressed in Section 4, where we leverage the information geometry of the considered statistical model in order to propose a statistically efficient on-line estimation process.

3. Robust GLRT with Kronecker product structure

The wavelet transform in [1] consists in combining the polarization channels to reflect the spectro-angular diversity of the data. This combination implies a Kronecker product structure of the covariance matrix, in which the size of one of the two matrices will be 3 (number of polarization channels), while the dimension of the other will be equal to the number of frequency times the angular intervals that are considered. Hence, the covariance matrix $\boldsymbol{\Sigma}$ can be assumed to be structured as $\boldsymbol{\Sigma} = \boldsymbol{A} \otimes \boldsymbol{B}$, with $\boldsymbol{A} \in \mathcal{H}_a^{++}$ and $\boldsymbol{B} \in \mathcal{H}_b^{++}$, where $a = 3$ and $b = p/3$[1]. To avoid any scaling ambiguity on this decomposition and the scaling by the textures, an arbitrary normalization on $\boldsymbol{A}$ and $\boldsymbol{B}$ is needed. We rely on the unit determinant for both matrices $\boldsymbol{A}$ and

[1]This remains valid whatever a and b.

$\boldsymbol{B}$, which, as noted in [28, 23], is particularly advantageous from a geometrical point of view. In this case, we have $\boldsymbol{A} \in s\mathcal{H}_a^{++} = \{\boldsymbol{M} \in \mathcal{H}_a^{++} : |\boldsymbol{M}| = 1\}$ and $\boldsymbol{B} \in s\mathcal{H}_b^{++} = \{\boldsymbol{M} \in \mathcal{H}_b^{++} : |\boldsymbol{M}| = 1\}$. Following from the SG model of Section 2, we have that each sample of a patch is distributed according to $\boldsymbol{x}_i \sim \mathcal{CN}(\boldsymbol{0}, \tau_i \boldsymbol{A} \otimes \boldsymbol{B})$. The likelihood of a pixel patch at time t is then directly transposed from (1), and also denoted $\mathcal{L}(\{\boldsymbol{x}_i^{(t)}\}_{i\in[\![1,n]\!]}, \theta^{(t)})$, in which the parameter of interest $\theta^{(t)}$ now denotes $\theta^{(t)} = \{\boldsymbol{A}^{(t)}, \boldsymbol{B}^{(t)}, \boldsymbol{\tau}^{(t)}\}$, which lies in the product manifold $\mathcal{M} = s\mathcal{H}_a^{++} \times s\mathcal{H}_b^{++} \times \mathbb{R}_{++}^n$.

The following proposition then adapts the GLRT of (3) to the case where the covariance matrix has such Kronecker product structure.

Proposition 1. *The GLRT for the problem of detection* (2) *when* $\theta^{(t)} = \{\boldsymbol{A}^{(t)}, \boldsymbol{B}^{(t)}, \boldsymbol{\tau}^{(t)}\}$ *and* T *images is:*

$$\log(\hat{\Lambda}_{K-SG}^{(T)}) = \mathcal{L}^{\mathrm{H}_1}(\{\hat{\theta}^{(t)}\}_{t=1}^T) - \mathcal{L}^{\mathrm{H}_0}(\hat{\theta}^{(0)}) \tag{6}$$

where the log-likelihoods of the whole data under H_0 *and* H_1 *are expressed as:*

$$\begin{aligned}
\mathcal{L}^{\mathrm{H}_0}(\theta^{(0)}) &= -nT\log|\boldsymbol{A}^{(0)} \otimes \boldsymbol{B}^{(0)}| - Tp\textstyle\sum_{i=1}^n \log(\tau_i^{(0)}) \\
&\quad - \textstyle\sum_{i=1}^n \sum_{t=1}^T \frac{\boldsymbol{x}_i^{(t)\,H}(\boldsymbol{A}^{(0)}\otimes\boldsymbol{B}^{(0)})^{-1}\boldsymbol{x}_i^{(t)}}{\tau_i^{(0)}} \\
\mathcal{L}^{\mathrm{H}_1}(\{\theta^{(t)}\}_{t=1}^T) &= -n\textstyle\sum_{t=1}^T \log|\boldsymbol{A}^{(t)} \otimes \boldsymbol{B}^{(t)}| - p\sum_{i=1}^n \sum_{t=1}^T \log(\tau_i^{(t)}) \\
&\quad - \textstyle\sum_{i=1}^n \sum_{t=1}^T \frac{\boldsymbol{x}_i^{(t)\,H}(\boldsymbol{A}^{(t)}\otimes\boldsymbol{B}^{(t)})^{-1}\boldsymbol{x}_i^{(t)}}{\tau_i^{(t)}},
\end{aligned} \tag{7}$$

and where $\hat{\theta}^{(0)} = \{\hat{\boldsymbol{A}}^{(0)}, \hat{\boldsymbol{B}}^{(0)}, \hat{\boldsymbol{\tau}}^{(0)}\}$ *stands for the maximum likelihood estimator (MLE) of the parameters under* H_0*, and* $\hat{\theta}^{(t)} = \{\hat{\boldsymbol{A}}^{(t)}, \hat{\boldsymbol{B}}^{(t)}, \hat{\boldsymbol{\tau}}^{(t)}\}$, $\forall t \in [\![1, T]\!]$ *for the MLE of the parameters under* H_1*. These MLEs are expressed respectively as*

$$\begin{aligned}
\hat{\boldsymbol{A}}^{(0)} &= \tfrac{1}{Tnb}\left(\textstyle\sum_{t=1}^T \sum_{i=1}^n \frac{\boldsymbol{M}_{i,t}^T(\hat{\boldsymbol{B}}^{(0)})^{-1*}\boldsymbol{M}_{i,t}^*}{\hat{\tau}_i^{(0)}}\right) \\
\hat{\boldsymbol{B}}^{(0)} &= \tfrac{1}{Tna}\left(\textstyle\sum_{t=1}^T \sum_{i=1}^n \frac{\boldsymbol{M}_{i,t}(\hat{\boldsymbol{A}}^{(0)})^{-1*}\boldsymbol{M}_{i,t}^H}{\hat{\tau}_i^{(0)}}\right)
\end{aligned} \quad \textbf{\textit{and}}\ \hat{\tau}_i^{(0)} = \sum_{t=1}^T \frac{\boldsymbol{x}_i^{(t)\,H}(\hat{\boldsymbol{A}}^{(0)} \otimes \hat{\boldsymbol{B}}^{(0)})^{-1}\boldsymbol{x}_i^{(t)}}{Tp} \tag{8}$$

and

$$\begin{aligned}
\hat{\boldsymbol{A}}^{(t)} &= \tfrac{1}{nb}\left(\textstyle\sum_{i=1}^n \frac{\boldsymbol{M}_{i,t}^T(\hat{\boldsymbol{B}}^{(t)})^{-1*}\boldsymbol{M}_{i,t}^*}{\hat{\tau}_i^{(t)}}\right) \\
\hat{\boldsymbol{B}}^{(t)} &= \tfrac{1}{na}\left(\textstyle\sum_{i=1}^n \frac{\boldsymbol{M}_{i,t}(\hat{\boldsymbol{A}}^{(t)})^{-1*}\boldsymbol{M}_{i,t}^H}{\hat{\tau}_i^{(t)}}\right)
\end{aligned} \quad \textbf{\textit{and}}\ \hat{\tau}_i^{(t)} = \frac{\boldsymbol{x}_i^{(t)\,H}(\hat{\boldsymbol{A}}^{(t)} \otimes \hat{\boldsymbol{B}}^{(t)})^{-1}\boldsymbol{x}_i^{(t)}}{p} \tag{9}$$

where $\boldsymbol{M}_{i,t} \in \mathbb{C}^{a\times b}$ *is matrix storing the entries of* $\boldsymbol{x}_i^{(t)}$*, such that* $\mathrm{vec}(\boldsymbol{M}_{i,t}) = \boldsymbol{x}_i^{(t)}$*. All these MLE estimators in* (8) *and* (9) *can be evaluated with a fixed-point algorithm, such as the one proposed in [20].*

Proof. The principle of the GLRT consists in computing the quantity:

$$\log(\hat{\Lambda}_{K-SG}^{(T)}) = \max_{\{\theta^{(t)}\}_{t=1}^{T}} \quad \mathcal{L}^{\mathrm{H}_1}(\{\theta^{(t)}\}_{t=1}^{T}) - \max_{\theta^{(0)}} \quad \mathcal{L}^{\mathrm{H}_0}(\theta^{(t)}). \tag{10}$$

Obtaining the maximal values in the above expression then boils down to computing the MLE under H_0 and H_1. This requires canceling the derivative of $\mathcal{L}^{\mathrm{H}_0}$ (resp. $\mathcal{L}^{\mathrm{H}_1}$) with respect to $\boldsymbol{A}^{(0)}$, $\boldsymbol{B}^{(0)}$, and $\boldsymbol{\tau}^{(0)}$ (resp. $\boldsymbol{A}^{(t)}$, $\boldsymbol{B}^{(t)}$, and $\boldsymbol{\tau}^{(t)}$ $\forall t$). The derivatives of the trace terms are obtained by using $\boldsymbol{x}^H(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{A}^{-T}\boldsymbol{M}^H\boldsymbol{B}^{-1}\boldsymbol{M})$ and the following derivative:

$$\frac{\partial \,\mathrm{tr}(\boldsymbol{O}\boldsymbol{X}^{-1}\boldsymbol{P})}{\partial \boldsymbol{X}} = -(\boldsymbol{X}^{-1}\boldsymbol{P}\boldsymbol{O}\boldsymbol{X}^{-1})^H, \tag{11}$$

which yields the following results:

$$\begin{array}{rcl} \frac{\partial \boldsymbol{x}^H(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}\boldsymbol{x}}{\partial \boldsymbol{A}} & = & -\boldsymbol{A}^{-1}\boldsymbol{M}^T(\boldsymbol{B}^{-1})^*\boldsymbol{M}^*\boldsymbol{A}^{-1} \\ \frac{\partial \boldsymbol{x}^H(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}\boldsymbol{x}}{\partial \boldsymbol{B}} & = & -\boldsymbol{B}^{-1}\boldsymbol{M}(\boldsymbol{A}^{-1})^*\boldsymbol{M}^H\boldsymbol{B}^{-1}. \end{array} \tag{12}$$

The derivatives of the log terms are obtained by using the relation $\log|\boldsymbol{A}\otimes\boldsymbol{B}| = b\log|\boldsymbol{A}| + a\log|\boldsymbol{B}|$, and $\frac{\partial \log|\boldsymbol{X}|}{\partial \boldsymbol{X}} = \boldsymbol{X}^{-H}$. After some matrix manipulations, we obtain the final results in (8) and (9). □

Remark: the expression of the MLEs under H_1 coincide with Tyler's M-estimator with Kronecker product structure obtained in [20] (slightly modified, as we deal with complex valued matrices). Those under H_0 are slightly different due to the assumed shared parameters between the different time index, and thus involve a summation over t.

Similarly to the unstructured GLRT in (3), the detector (6) cannot be recursively derived when a new patch $\mathcal{X}_{T+1} = \{\boldsymbol{x}_i^{(T+1)}\}_{i=1}^n$ is added to the image stack. Indeed, if we denote the observation sets $\mathcal{X}_{1:T} = \{\{\boldsymbol{x}_i^{(t)}\}_{i=1}^n\}_{t=1}^T$ and $\mathcal{X}_{1:T+1} = \{\{\boldsymbol{x}_i^{(T+1}\}_{i=1}^n\}_{t=1}^{T+1}$, it is easy to notice for the MLE of $\boldsymbol{A}$ under H_0 that:

$$\hat{\boldsymbol{A}}^{(0)}(\mathcal{X}_{1:T+1}) \neq \frac{T}{T+1}\hat{\boldsymbol{A}}^{(0)}(\mathcal{X}_{1:T}) + \frac{1}{T+1}\hat{\boldsymbol{A}}^{(0)}(\mathcal{X}_{T+1}) \tag{13}$$

due to the intricacy of the fixed-point solutions (the same is also true for $\boldsymbol{B}$ and so on $\boldsymbol{\tau}$). When processing large time series, re-computing the fixed point can then be computationally demanding for large dimensions p and T, which limits the practical implementation of the GLRTs. To overcome this issue, we propose in the next section to recursively update the parameters $\theta^{(0)}$, which allows for computing an online version of the detector (6). Such recursive algorithm, inspired by [23], will be obtained by leveraging the information geometry of the manifold $\mathcal{M}$ induced by the SG distribution.

4. Online Robust GLRT with Kronecker product structure

In this section, we first derive the Riemannian geometry of $\mathcal{M}$ equipped with the Fisher information metric induced by the likelihood (1). The exponential mapping which is needed in any Riemannian gradient descent algorithm is also given as well as the corresponding geodesic distance. The second part of this section is devoted to the online estimation of the parameters $\theta^{(0)}$ with a stochastic Riemannian gradient descent algorithm.

4.1. Information geometry of $\mathcal{M}$

In this subsection, we will omit the subscript $.^{(t)}$ since the results given in this section will be valid for any values of T. By definition, the tangent space of $\mathcal{M}$ at θ is $T_\theta\mathcal{M} = T_{\boldsymbol{A}}s\mathcal{H}_a^{++} \times T_{\boldsymbol{B}}s\mathcal{H}_b^{++} \times T_{\boldsymbol{\tau}}\mathbb{R}_{++}^n$. The tangent space $T_{\boldsymbol{\tau}}\mathbb{R}_{++}^n$ is identified to $\mathbb{R}^n$ and the tangent spaces $T_{\boldsymbol{A}}\mathcal{H}_a^{++}$ and $T_{\boldsymbol{B}}\mathcal{H}_b^{++}$ can be identified by [26]

$$\begin{array}{rcl} T_{\boldsymbol{A}}s\mathcal{H}_a^{++} & = & \{\boldsymbol{\xi_A} \in \mathcal{H}_a : \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi_A}) = 0\}. \\ T_{\boldsymbol{B}}s\mathcal{H}_b^{++} & = & \{\boldsymbol{\xi_B} \in \mathcal{H}_b : \operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\xi_B}) = 0\}. \end{array} \tag{14}$$

In the following, $\xi = (\boldsymbol{\xi_A}, \boldsymbol{\xi_B}, \boldsymbol{\xi_\tau})$ and $\eta = (\boldsymbol{\eta_A}, \boldsymbol{\eta_B}, \boldsymbol{\eta_\tau})$ denote two elements from $T_\theta\mathcal{M}$. The Fisher information metric on $\mathcal{M}$ induced by the scaled Gaussian distribution is provided in Proposition 2.

Proposition 2 (Fisher information metric)**.** *Given $\theta \in \mathcal{M}$, ξ and $\eta \in T_\theta\mathcal{M}$, the Fisher information metric on $\mathcal{M}$ induced by the likelihood* (1) *is*

$$\langle \xi, \eta \rangle_\theta = \frac{b}{p}\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi_A}\boldsymbol{A}^{-1}\boldsymbol{\eta_A}) + \frac{a}{p}\operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\xi_B}\boldsymbol{B}^{-1}\boldsymbol{\eta_B}) + \frac{1}{n}(\boldsymbol{\xi_\tau} \odot \boldsymbol{\tau}^{\odot^{-1}})^T(\boldsymbol{\eta_\tau} \odot \boldsymbol{\tau}^{\odot^{-1}})$$

Proof. We first introduce the mapping $\varphi_i : \mathcal{M} \to \mathcal{H}_p^{++}$ as:

$$\varphi_i(\theta^{(t)}) = \tau_i^{(t)} \boldsymbol{A}^{(t)} \otimes \boldsymbol{B}^{(t)}, \tag{15}$$

such that $\boldsymbol{x}_i^{(t)} \sim \mathcal{CN}(\boldsymbol{0}, \varphi_i(\theta^{(t)}))$. From [29, Proposition 7], we have

$$\langle \xi, \eta \rangle_\theta = \langle \mathrm{D}\, \varphi_i(\theta)[\xi], \mathrm{D}\, \varphi(\theta)[\eta] \rangle_{\varphi(\theta)}^{\mathcal{H}_p^{++}},$$

where $\langle \cdot, \cdot \rangle_{\cdot}^{\mathcal{H}_p^{++}}$ is the Fisher metric of the Gaussian distribution on $\mathcal{H}_p^{++}$. This metric is, e.g., obtained in [25, 26], and is defined by:

$$\langle \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} = \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_2) \tag{16}$$

The directional derivative of φ at θ in the direction ξ is obtained by $\phi_i(\theta + \xi) - \phi_i(\theta)$ and the result is then

$$\mathrm{D}\, \varphi(\theta)[\xi] = \tau_i \boldsymbol{A} \otimes \boldsymbol{\xi}_{\boldsymbol{B}} + \tau_i \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\xi}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B}.$$

The result is obtained by plugging $\mathrm{D}\, \varphi(\theta)[\xi]$ into $\langle \cdot, \cdot \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}}$, we have:

$$\begin{aligned}
\langle \xi, \eta \rangle_\theta = \sum_{i=1}^{n} \langle & \tau_i \boldsymbol{A} \otimes \boldsymbol{\xi}_{\boldsymbol{B}} + \tau_i \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\xi}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B}, \\
& \tau_i \boldsymbol{A} \otimes \boldsymbol{\eta}_{\boldsymbol{B}} + \tau_i \boldsymbol{\eta}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\eta}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} \\
= \mathrm{tr} \big\{ & (\tau_i \boldsymbol{A} \otimes \boldsymbol{B})^{-1} (\tau_i \boldsymbol{A} \otimes \boldsymbol{\xi}_{\boldsymbol{B}} + \tau_i \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\xi}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B}) \\
& (\tau_i \boldsymbol{A} \otimes \boldsymbol{B})^{-1} (\tau_i \boldsymbol{A} \otimes \boldsymbol{\eta}_{\boldsymbol{B}} + \tau_i \boldsymbol{\eta}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\eta}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B}) \big\}
\end{aligned}$$

Let us compute

$$\begin{aligned}
(\tau_i \boldsymbol{A} \otimes \boldsymbol{B})^{-1} & (\tau_i \boldsymbol{A} \otimes \boldsymbol{\xi}_{\boldsymbol{B}} + \tau_i \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{B} + \boldsymbol{\xi}_{\tau_i} \boldsymbol{A} \otimes \boldsymbol{B}) \\
& = \boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}^{-1} \boldsymbol{\xi}_{\boldsymbol{B}} + \frac{\boldsymbol{\xi}_{\tau_i}}{\tau_i} \boldsymbol{I}.
\end{aligned}$$

By exploiting the relations $(\boldsymbol{M} \otimes \boldsymbol{N})(\boldsymbol{O} \otimes \boldsymbol{P}) = \boldsymbol{MO} \otimes \boldsymbol{NP}$, $\mathrm{tr}(\boldsymbol{M} \otimes \boldsymbol{N}) = \mathrm{tr}(\boldsymbol{M})\, \mathrm{tr}(\boldsymbol{N})$ and $\mathrm{tr}(\boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}}) = \mathrm{tr}(\boldsymbol{A}^{-1} \boldsymbol{\eta}_{\boldsymbol{A}}) = 0$ (and same property for matrix $\boldsymbol{B}$), we obtain the announced result. □

Remark: we notice that the determinant constraint is the best choice with respect to information geometry since it allows to obtain a separable metric. This is not the case when the trace constraint is used instead.

One can notice that the metric of Proposition 2 is separable into two scaled Riemannian affine invariant metrics on $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and a last term on $\boldsymbol{\tau}$. The geometry of a product manifold resulting from such a separable metric is simply obtained by combining the geometries corresponding to each component. In particular, the Riemannian exponential mapping at $\theta \in \mathcal{M}$ is defined for $\xi \in T_\theta\mathcal{M}$ as

$$\exp_\theta^{\mathcal{M}}(\xi) = \left(\boldsymbol{A}\exp(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}), \boldsymbol{B}\exp(\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}}), \boldsymbol{\tau} \odot \exp(\boldsymbol{\tau}^{\odot^{-1}} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})\right). \quad (17)$$

This exponential mapping will be very useful when it comes to optimizing a cost function on the manifold. As for the exponential mapping, it is easy to derive the geodesic distance since the metric is completely separable, and we have:

$$\delta^2_{\mathcal{M}}(\theta_0, \theta_1) = \frac{b}{p}\delta^2_{\mathcal{H}_a^{++}}(\boldsymbol{A}_0, \boldsymbol{A}_1) + \frac{a}{p}\delta^2_{\mathcal{H}_b^{++}}(\boldsymbol{B}_0, \boldsymbol{B}_1) + \frac{1}{n}\delta^2_{\mathbb{R}^n_{++}}(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1), \quad (18)$$

where $\delta^2_{\mathcal{H}_p^{++}}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \| \operatorname{logm}(\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1/2})\|_2^2$ and $\delta^2_{\mathbb{R}^n_{++}}(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1) = \| \log(\boldsymbol{\tau}_0^{-1} \odot \boldsymbol{\tau}_1)\|_2^2$.

Thanks to this geometry, we are now able to derive an online change detection when the covariance matrix possesses a Kronecker structure and the data follows a SG distribution.

4.2. Online estimation of $\boldsymbol{A}_0$, $\boldsymbol{B}_0$ and $\boldsymbol{\tau}_0$

Given a new pixel patch $\mathcal{X}_{T+1} = \{\boldsymbol{x}_i^{(T+1)}\}_{i\in[\![1,n]\!]}$ in the time-series, the MLE of the parameter $\theta^{(0)} = \{\boldsymbol{A}^{(0)}, \boldsymbol{B}^{(0)}, \boldsymbol{\tau}^{(0)}\}$ is expressed as the fixed point (8), that depends on the whole data $\mathcal{X}_{1:T+1}$. In order to evaluate the proposed GLRT in a lightweight streaming fashion, our goal is to derive a recursive algorithm that provides an estimate of $\theta^{(0)}$ at time $T+1$, denoted $\hat{\theta}^{(0)}_{\mathcal{X}_{T+1}}$, given solely the previous estimate $\hat{\theta}^{(0)}_{\mathcal{X}_{1:T}}$ and the incoming data $\mathcal{X}_{T+1}$. Though many options can be envisioned to obtained such recursive algorithm, the interest of leveraging the stochastic natural gradient descent [22, 29] is twofold:

- By leveraging the Riemannian gradient associated with the Fisher information metric (information geometry), a theoretically optimal gradient

step can easily be obtained (hence, no line search will be involved in the update). Furthermore, this step only depends on intrinsic dimensions (a, b, n) of the estimation problem.

- The corresponding stochastic Riemannian gradient descent ensures statistical efficiency, i.e., that the recursive estimate $\hat{\theta}^{(0)}_{\mathcal{X}_{1:T+1}}$ tends to the MLE of $\theta^{(0)}$ under H_0 when $T \to \infty$, so the corresponding online GLRT is also asymptotically consistent.

An iteration of the proposed stochastic natural gradient descent [22, 29] is expressed as:

$$\hat{\theta}^{(0)}_{\mathcal{X}_{1:T+1}} = \exp^{\mathcal{M}}_{\hat{\theta}^{(0)}_{\mathcal{X}_{1:T}}} \left(-\frac{\alpha_0}{T} \operatorname{grad} \mathcal{L}^{(T+1)}(\hat{\theta}^{(0)}_{\mathcal{X}_{1:T}}) \right) \tag{19}$$

where $\exp^{\mathcal{M}}_{\cdot}$ is given in (17), and where $\operatorname{grad} \mathcal{L}^{(T+1)}(\hat{\theta}^{(0)}_{\mathcal{X}_{1:T}})$ denotes the Riemannain gradient of the likelihood of the data $\mathcal{X}_{T+1}$ evaluated at point $\hat{\theta}^{(0)}_{\mathcal{X}_{1:T}}$. In order to apply this algorithm to the problem of interest, it remains to compute the Riemannian gradient of $\mathcal{L}^{(T+1)}$ according to the metric of Proposition 2. This Riemannian gradient is given in Proposition 3.

Proposition 3 (Riemannian gradient)**.** *The Riemannian gradient of $\mathcal{L}^{(T+1)}$ at $\theta \in \mathcal{M}$ according to the metric of Proposition 2 is given by*

$$\operatorname{grad} \mathcal{L}^{(T+1)}(\theta) = \left(-\sum_{i=1}^{n} \frac{1}{b\tau_i} P_{\boldsymbol{A}}(\boldsymbol{M}_i^T \boldsymbol{B}^{-T} \boldsymbol{M}_i^*) \right.$$
$$\left. , -\sum_{i=1}^{n} \frac{1}{a\tau_i} P_{\boldsymbol{B}}(\boldsymbol{M}_i \boldsymbol{A}^{-T} \boldsymbol{M}_i^H), n(\boldsymbol{q}^{(T+1)} - p\boldsymbol{\tau}) \right),$$

where $q_i(\boldsymbol{A}, \boldsymbol{B}) = \boldsymbol{x}_i^{(T+1)\,H} (\boldsymbol{A} \otimes \boldsymbol{B})^{-1} \boldsymbol{x}_i^{(T+1)}$ and $\boldsymbol{q}^{(T+1)} = (q_1(\boldsymbol{A}, \boldsymbol{B}), \ldots, q_n(\boldsymbol{A}, \boldsymbol{B}))^T$. The matrix $P_{\boldsymbol{A}} : \mathcal{H}_a \to T_{\boldsymbol{A}} s\mathcal{H}_a^{++}$ is the orthogonal projection map such that

$$P_{\boldsymbol{A}}(\boldsymbol{\xi}_{\boldsymbol{A}}) = \boldsymbol{\xi}_{\boldsymbol{A}} - \frac{\operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}})}{a} \boldsymbol{A}.$$

and the matrix $P_{\boldsymbol{B}}(\boldsymbol{\xi}_{\boldsymbol{B}})$ follows from the fame expression.

Proof. The likelihood of the new dataset $\{\boldsymbol{x}_i^{(T+1)}\}_{i \in [\![1,n]\!]}$ is

$$\mathcal{L}^{(T+1)}(\theta^{(T+1)}) = \sum_{i=1}^{n} \ell_i(\theta^{(T+1)}), \tag{20}$$

where

$$\ell_i(\theta) = -p\log(\tau_i) - \frac{q_i(\boldsymbol{A},\boldsymbol{B})}{\tau_i},$$

with $q_i(\boldsymbol{A},\boldsymbol{B}) = \boldsymbol{x}_i^{(T+1)\,H}(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}\boldsymbol{x}_i^{(T+1)}$. Notice that the log-determinant terms in the likelihood vanish because we consider matrices of determinant one. The Riemannian gradient is defined according to the metric through the relation:

$$\langle \operatorname{grad}\mathcal{L}^{(T+1)}(\theta), \boldsymbol{\xi}\rangle_\theta^{\mathcal{M}} = \mathrm{D}\,\mathcal{L}^{(T+1)}(\theta)[\boldsymbol{\xi}]. \tag{21}$$

Thus we need the directional derivative of $\mathcal{L}^{(T+1)}(\theta)$:

$$\begin{aligned}\mathrm{D}\,\mathcal{L}^{(T+1)}(\theta)[\boldsymbol{\xi}] &= -\sum_{i=1}^n \left(\tfrac{p\boldsymbol{\xi}_{\tau_i}}{\tau_i} + \tfrac{1}{\tau_i^2}q_i(\boldsymbol{A},\boldsymbol{B})\boldsymbol{\xi}_{\tau_i}\right) - \sum_{i=1}^n \tfrac{\mathrm{D}\,q_i(\boldsymbol{A},\boldsymbol{B})[\xi]}{\tau_i} \\ &= \tfrac{1}{n}\langle n(\boldsymbol{q}^{(t)} - p\boldsymbol{\tau}), \boldsymbol{\xi}_{\boldsymbol{\tau}}\rangle_\tau^{\mathbb{R}^n_{++}} - \sum_{i=1}^n \tfrac{\mathrm{D}\,q_i(\boldsymbol{A},\boldsymbol{B})[\xi]}{\tau_i}\end{aligned} \tag{22}$$

where $\langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}}\rangle_\tau^{\mathbb{R}^n_{++}} = (\boldsymbol{\xi}_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{\odot^{-1}})^T(\boldsymbol{\eta}_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{\odot^{-1}})$. The second term can be obtained thanks to the following result:

$$\begin{array}{cc} g(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X}\boldsymbol{x}\boldsymbol{x}^H), & \mathrm{D}\,g(\boldsymbol{X})[\boldsymbol{\xi}] = \operatorname{tr}(\boldsymbol{\xi}\boldsymbol{x}\boldsymbol{x}^H) \\ g(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X}^{-1}\boldsymbol{x}\boldsymbol{x}^H), & \mathrm{D}\,g(\boldsymbol{X})[\boldsymbol{\xi}] = \operatorname{tr}(\boldsymbol{X}^{-1}\boldsymbol{\xi}\boldsymbol{X}^{-1}\boldsymbol{x}\boldsymbol{x}^H) \end{array} \tag{23}$$

which leads to

$$\begin{aligned}\mathrm{D}\,q_i(\boldsymbol{A},\boldsymbol{B})[\xi] &= -\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}\boldsymbol{A}^{-1}\boldsymbol{M}_i^T\boldsymbol{B}^{-T}\boldsymbol{M}_i^*) \\ &\quad -\operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}}\boldsymbol{B}^{-1}\boldsymbol{M}_i\boldsymbol{A}^{-T}\boldsymbol{M}_i^*) \\ &= -\langle \boldsymbol{M}_i^T\boldsymbol{B}^{-T}\boldsymbol{M}_i^*, \boldsymbol{\xi}_{\boldsymbol{A}}\rangle_{\boldsymbol{A}}^{\mathcal{H}_a^{++}} - \langle \boldsymbol{M}_i\boldsymbol{A}^{-T}\boldsymbol{M}_i^H, \boldsymbol{\xi}_{\boldsymbol{B}}\rangle_{\boldsymbol{B}}^{\mathcal{H}_b^{++}}\end{aligned} \tag{24}$$

Finally, the result of the Riemannian gradient according to the metric of Proposition 2 is obtained through identification by (21) and projection onto the tangent spaces. □

Thanks to this proposition, the online GLRT is finally computed by plugging the estimates updates from (19) in the detector (6).

To evaluate the quality of the estimation of parameters $\boldsymbol{A}_0$, $\boldsymbol{B}_0$ and $\boldsymbol{\tau}_0$, it is interesting to derive the intrinsic Cramér-Rao bound (ICRB) [25, 29] and the corresponding inequality. This is the purpose of the next proposition.

Proposition 4 (ICRB)**.** *Given an unbiased estimator $\hat{\theta}^{(T)}$ of $\hat{\theta}^{(T)}_{\mathcal{X}_{1:T}}$ corresponding to a MITS with T data, the ICRB corresponding to the error mea-*

sured with the Riemannian distance (18) *and the metric given in the proposition* (2) *is*

$$\mathbb{E}[\delta^2_{\mathcal{M}}(\theta^{(T)}, \hat{\theta}^{(T)}_{\mathcal{X}_{1:T}})] \leq \frac{(a^2-1)+(b^2-1)+n}{Tpn}$$

Proof. Since the distance is obtained trough the Fisher information metric, the ICRB for the problem of estimation of $\theta^{(T)}$ could be easily deduced from the approach of [25, 29] and is the ratio between the dimension of the problems and the number of parameters to estimate. □

Thanks to the fact that the metric is separable, it is then easy to obtain the following inequalities for each component of $\theta^{(T)}$:

$$\begin{array}{lcl} \mathbb{E}[\delta^2_{\mathcal{H}_a^{++}}(\hat{\boldsymbol{A}}^{(T)}_{\mathcal{X}_{1:T}}, \boldsymbol{A}^{(T)})] & \leq & \frac{(a^2-1)}{bTn}, \\ \mathbb{E}[\delta^2_{\mathcal{H}_b^{++}}(\hat{\boldsymbol{B}}^{(T)}_{\mathcal{X}_{1:T}}, \boldsymbol{B}^{(T)})] & \leq & \frac{(b^2-1)}{aTn}, \\ \mathbb{E}[\delta^2_{\mathbb{R}^n_{++}}(\hat{\boldsymbol{\tau}}^{(T)}_{\mathcal{X}_{1:T}}, \boldsymbol{\tau}^{(T)})] & \leq & \frac{1}{Tp}. \end{array} \tag{25}$$

5. Numerical experiments

In this section, we first show the performances in terms of MSE of the online estimator proposed in 4. Next, we will show the interest of this online approach to the problem of change detection with simulated data as well as with an experiment composed of real data.

5.1. Estimation performance

In this section, we compare the performance of the maximum likelihood estimator obtained with a usual Riemannian optimization algorithm and of the one obtained with the on-line procedure of (19). In order to do so, simulated data drawn from the multivariate K-distribution with $\nu = 1$ for the shape parameter are generated with a Kronecker product structured covariance. To build a $p \times p$ ($p = 12$) covariance matrix, we compute

$$\begin{gathered} \boldsymbol{\Sigma} = \boldsymbol{A} \otimes \boldsymbol{B}, \\ \boldsymbol{A} = \boldsymbol{U_A}\boldsymbol{\Lambda_A}\boldsymbol{U_A}^T, \qquad \boldsymbol{B} = \boldsymbol{U_B}\boldsymbol{\Lambda_B}\boldsymbol{U_B}^T, \end{gathered} \tag{26}$$

where $a = 4$ and $b = 3$,

- $\boldsymbol{U_A}$ and $\boldsymbol{U_B}$ are random orthogonal matrices,

- $\boldsymbol{\Lambda_A}$ and $\boldsymbol{\Lambda_B}$ are diagonal matrices whose minimal and maximal elements are $1/\sqrt{c}$ and $\sqrt{c}$ ($c = 10$ is the condition number with respect to inversion); their other elements are randomly drawn from the uniform distribution between $1/\sqrt{c}$ and $\sqrt{c}$; the determinant of $\boldsymbol{\Lambda_A}$ is then normalized.

The number of samples is fixed to $n = 8$ since the conclusions keep the same for larger n. The number of frames T varies from 1 to 1000. 1000 trials are used to estimate the MSEs.

For this experiment, we consider the following estimators:

- the classical maximum-likelihood estimator obtained with Riemannian gradient descent (GD). Optimization for this estimator is performed with pymanopt toolbox [30, 31].

- the online version obtained through stochastic gradient descent (SGD) of (19).

The online algorithm is initialized with $\theta_0 = (\boldsymbol{A}^{(1)}, \boldsymbol{B}^{(1)}, \tau^{(1)})$ obtained with the GD when $T = 1$.

In order to measure the performance of the estimators, the distances defined in (18), $\delta^2_{\mathcal{H}_a^{++}}$, $\delta^2_{\mathcal{H}_a^{++}}$ and $\delta^2_{\mathbb{R}^n_{++}}$, will be used. Moreover, the ICRBs for each term given in (25) are also computed.

The results are presented in Figure 2 for the MSEs of $\boldsymbol{A}$ and $\boldsymbol{B}$ and in 3 for the MSE of $\boldsymbol{\tau}$. First, we notice that the on-line version converges to the classical estimate which is expected as stated in [22]. We notice that for a value of $T = 100$ the results of both algorithms become very close which is clearly interesting in terms of practical interest in applications like change detection.

5.2. Performance in change detection with simulated data

In this section, we compare the ROC plots of the change detection detector of (6) in the offline and online versions. We hope that for a large value of T, the performance become close. Like in the previous subsection, $p = 12$ with $a = 4$ and $b = 3$. We also simulate a K-distribution for the data. But, the covariance matrices are differently built to be better fitted with the change detection application. In both hypothesis, $\boldsymbol{A}$ and $\boldsymbol{B}$ are Toeplitz matrices of correlation coefficient ρ_0 and ρ_1 (for $\mathbf{A}$: $\rho_0 = 0.3 + 0.7j$,

$\rho_1 = 0.3 + 0.5j$. For $\mathbf{B}$: $\rho_0 = 0.3 + 0.6j$, $\rho_1 = 0.4 + 0.5j$). For the H_1 hypothesis, the change is put at the middle of the time series, i.e. at $T/2$. The number of samples n is 13. To show the interest to take into account of the Kronecker structure in the change detection algorithm, we compare our two versions of the Kronecker change detection to the one proposed in [14] and given in (3). To resume, we test 4 change detectors, $\hat{\Lambda}_{SG}^{(T)}$ and $\hat{\Lambda}_{K-SG}^{(T)}$, and their corresponding on-line versions, $\hat{\Lambda}_{SG-O}^{(T)}$ and $\hat{\Lambda}_{K-SG-O}^{(T)}$. For the offline versions, the number of images T is equal to 50 whereas the online versions are built from T varying from 2 to 50. All ROC plots are estimated by using 5000 trials.

Figure 4 shows the ROC plots when the data follows a Gaussian distribution. Figures 5 and 6 show the same plots but when the data follow a K distribution with $\nu = 100$ and $\nu = 1$ as shape parameters respectively. We notice that all the Kronecker versions outperform the classical ones which is expected by the structure of the covariance matrix. For the heterogeneous hypothesis ($\nu = 1$), we get a very good result, since online versions perform similarly to offline detection when $T = 50$. Conversely, we have the surprising result that online versions have a slower convergence speed for Gaussian or quasi-Gaussian data. Nevertheless, we can conclude that the detection results of the online detectors improve with T and converge towards the offline result, particularly when the Kronecker structure is used.

5.3. Performance in change detection with real data

The SAR ITS data set is taken from UAVSAR (courtesy of NASA/JPL-Caltech) and is referenced as SDelta_28518_02, Segment 1. The number of images available is $T = 68$. They show the evolution of a river delta in the USA with cycle of droughts and flood. Some strong scatterers appear and disappear over time. Since the image cover an extensive area of the delta, we crop it to an interesting part of size 200×200 pixels. The SAR images correspond to full-polarization data with a resolution of 1.67 m in range and 0.6 m in azimuth. Thanks to this high resolution property of the SAR images, the scatterers present in this scene exhibit an interesting spectro-angular behavior, each polarization of these images has been subjected to the wavelet transform presented in [5], allowing to obtain images of dimension $p = 12$. In this particular configuration, the full covariance matrices show an inherent Kronecker structure $\mathbf{A} \otimes \mathbf{B}$. The matrix $\mathbf{A} \in \mathbb{C}^{a \times a}$ corresponds to the spectro-angular property ($a = 4$) although $\mathbf{B} \in \mathbb{C}^{b \times b}$ is linked to the polarization ($b = 3$).

As in the previous section, we test 4 change detectors, $\hat{\Lambda}_{SG}^{(T)}$ and $\hat{\Lambda}_{K-SG}^{(T)}$, and their corresponding on-line versions, $\hat{\Lambda}_{SG-O}^{(T)}$ and $\hat{\Lambda}_{K-SG-O}^{(T)}$. Figure 7 shows the outputs of the four detectors for the complete time series. Since we do not have any ground truths we are not able to conclude if the Kronecker versions are better than the classical ones. But we are interested in the similarity between the online and offline versions. For $T = 68$, it is difficult to conclude since both results are quite different. We propose then to repeat the time series in order to have a larger one. Figure 8 shows the result when the number of repetition is 5 ($T = 340$). For this time series, we find better similarity between the online and corresponding offline versions, especially when the Kronecker structure is used in the change detector. This behavior is better illustrated in Figure 9 when the number of repetitions is 10 ($T = 680$).

6. Conclusion

In this paper, we developed an online version of a change detection algorithm that tests the difference in a series of a structured Kronecker covariance. The data are assumed to follow a SG distribution. This model is well adapted to SAR-MITS data. The development of the online version was based on the information geometry induced by this structure and the statistical assumption. The algorithm was mainly based on the online estimation of the covariance matrix under the null hypothesis. This step was done with a Riemannian gradient and we also provided the corresponding ICRB. Results on both simulated and real data showed the good results of this approach and its practical interest to analyze large image times series.

References

[1] A. Mian, J.-P. Ovarlez, A. M. Atto, G. Ginolhac, Design of new wavelet packets adapted to high-resolution SAR images with an application to target detection, IEEE Transactions on Geoscience and Remote Sensing 57 (6) (2019) 3919–3932.

[2] M. Hussain, D. Chen, A. Cheng, H. Wei, D. Stanley, Change detection from remotely sensed images: From pixel-based to object-based approaches, ISPRS Journal of Photogrammetry and Remote Sensing 80 (2013) 91–106. doi:https://doi.org/10.1016/j.isprsjprs.2013.03.006.
URL https://www.sciencedirect.com/science/article/pii/S0924271613000804

[3] A. Mian, G. Ginolhac, J. P. Ovarlez, A. Breloy, F. Pascal, An overview of covariance-based change detection methodologies in multivariate sar image time series, Chapter 3 in Change Detection and Image Time Ser. Anal., Unsupervised Methods, Vol. 1, ISTE/Wiley Encyclopedia of Sci.-Remote Sens. Imagery (2021).

[4] Y. Bazi, L. Bruzzone, F. Melgani, An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images, IEEE Transactions on Geoscience and Remote Sensing 43 (4) (2005) 874–887. doi:10.1109/TGRS.2004.842441.

[5] D. Ratha, S. De, T. Celik, A. Bhattacharya, Change detection in polarimetric sar images using a geodesic distance between scattering mechanisms, IEEE Geoscience and Remote Sensing Letters 14 (7) (2017) 1066–1070. doi:10.1109/LGRS.2017.2696158.

[6] A. D. C. Nascimento, A. C. Frery, R. J. Cintra, Detecting changes in fully polarimetric sar imagery with statistical information theory, IEEE Transactions on Geoscience and Remote Sensing 57 (3) (2019) 1380–1392. doi:10.1109/TGRS.2018.2866367.

[7] K. Conradsen, A. Nielsen, J. Schou, H. Skriver, A test statistic in the complex wishart distribution and its application to change detection in polarimetric sar data, IEEE Transactions on Geoscience and Remote Sensing 41 (1) (2003) 4–19. doi:10.1109/TGRS.2002.808066.

[8] D. Ciuonzo, V. Carotenuto, A. De Maio, On multiple covariance equality testing with application to sar change detection, IEEE Transactions on Signal Processing 65 (19) (2017) 5078–5091. doi:10.1109/TSP.2017.2712124.

[9] J. Billingsley, Ground clutter measurements for surface-sited radar, Tech. Rep. 780, MIT (February 1993).

[10] M. Greco, F. Gini, M. Rangaswamy, Statistical analysis of measured polarimetric clutter data at different range resolutions, Radar, Sonar and Navigation, IEE Proceedings - 153 (6) (2006) 473–481.

[11] E. Ollila, D. E. Tyler, V. Koivunen, H. V. Poor, Complex elliptically symmetric distributions: Survey, new results and applications, IEEE Transactions on signal processing 60 (11) (2012) 5597–5625.

[12] D. E. Tyler, A distribution-free M-estimator of multivariate scatter, The Annals of Statistics 15 (1) (1987) 234–251.

[13] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, P. Larzabal, Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis, IEEE Transactions on Signal Processing 56 (1) (2008) 34–48.

[14] A. Mian, G. Ginolhac, J. Ovarlez, A. M. Atto, New robust statistics for change detection in time series of multivariate SAR images, IEEE Transactions on Signal Processing 67 (2) (2019) 520–534. doi:10.1109/TSP.2018.2883011.

[15] R. B. Abdallah, A. Mian, A. Breloy, A. Taylor, M. N. El Korso, D. Lautru, Detection methods based on structured covariance matrices for multivariate sar images processing, IEEE Geoscience and Remote Sensing Letters 16 (7) (2019) 1160–1164.

[16] A. Mian, A. Collas, A. Breloy, G. Ginolhac, J.-P. Ovarlez, Robust low-rank change detection for multivariate SAR image time series, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020) 3545–3556. doi:10.1109/JSTARS.2020.2999615.

[17] K. Werner, M. Jansson, P. Stoica, On estimation of covariance matrices with Kronecker product structure, IEEE Transactions on Signal Processing 56 (2) (2008) 478–491.

[18] M. S. Srivastava, T. von Rosen, D. Von Rosen, Models with a Kronecker product covariance structure: estimation and testing, Mathematical Methods of Statistics 17 (4) (2008) 357–370.

[19] A. Wiesel, Geodesic convexity and covariance estimation, IEEE transactions on signal processing 60 (12) (2012) 6182–6189.

[20] Y. Sun, P. Babu, D. P. Palomar, Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions, IEEE Transactions on Signal Processing 64 (14) (2016) 3576–3590.

[21] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2009.

[22] J. Zhou, S. Said, Fast, asymptotically efficient, recursive estimation in a Riemannian manifold, Entropy 21 (10) (2019) 1021.

[23] F. Bouchard, A. Mian, J. Zhou, S. Said, G. Ginolhac, Y. Berthoumieu, Riemannian geometry for compound Gaussian distributions: application to recursive change detection, Signal Processing 176 (2020) 107716.

[24] F. Bouchard, A. Breloy, A. Mian, G. Ginolhac, On-line kronecker product structured covariance estimation with riemannian geometry for t-distributed data, in: 2021 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 856–859. doi:10.23919/EUSIPCO54536.2021.9616101.

[25] S. Smith, Covariance, subspace, and intrinsic Cramér-Rao bounds, IEEE Transactions on Signal Processing 53 (5) (2005) 1610–1630. doi:10.1109/TSP.2005.845428.

[26] A. Breloy, G. Ginolhac, A. Renaux, F. Bouchard, Intrinsic Cramèr – Rao bounds for scatter and shape matrices estimation in CES distributions, IEEE Signal Processing Letters 26 (2) (2019) 262–266. doi:10.1109/LSP.2018.2886700.

[27] E. Ollila, D. E. Tyler, Regularized M-estimators of scatter matrix, IEEE Transactions on Signal Processing 62 (22) (2014) 6059–6070.

[28] D. Paindaveine, A canonical definition of shape, Statist. Probability Lett. 78 (14) (2008) 2240–2247.

[29] F. Bouchard, A. Breloy, G. Ginolhac, A. Renaux, F. Pascal, A Riemannian framework for low-rank structured elliptical models, IEEE Transactions on Signal Processing 69 (2021) 1185–1199. doi:10.1109/TSP.2021.3054237.

[30] N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a Matlab toolbox for optimization on manifolds, Journal of Machine Learning Research 15 (42) (2014) 1455–1459.
URL `https://www.manopt.org`

[31] J. Townsend, N. Koep, S. Weichwald, Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation, Journal of Machine Learning Research 17 (137) (2016) 1–5.
URL `http://jmlr.org/papers/v17/16-177.html`

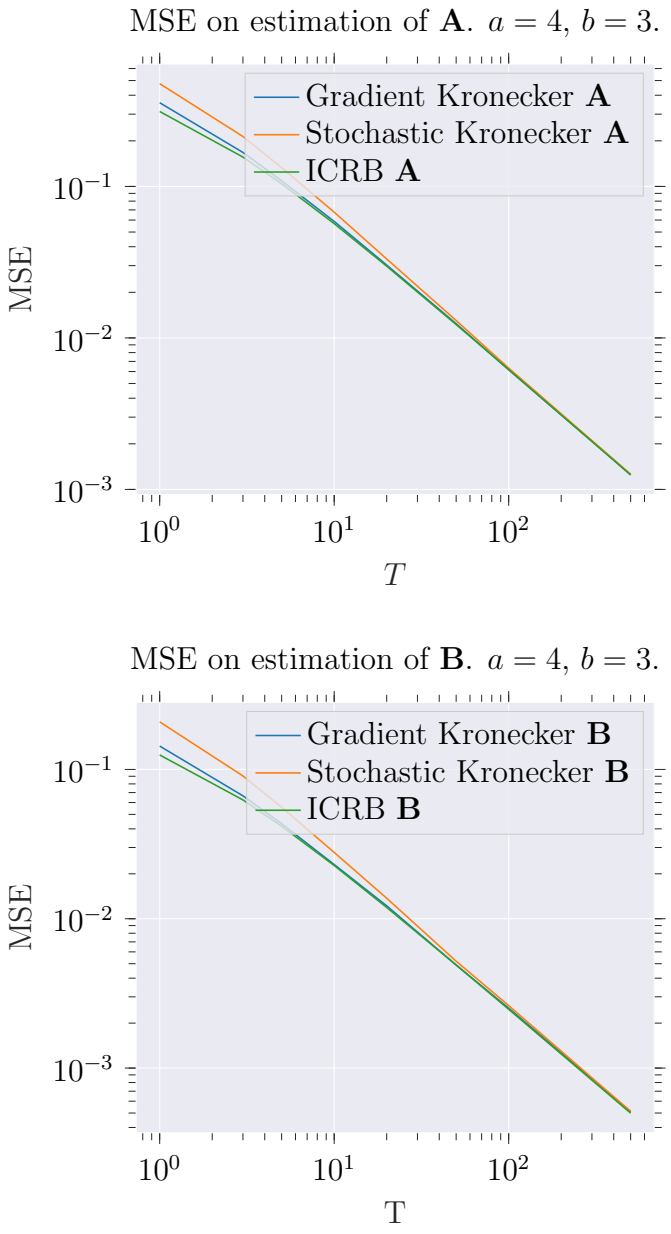


Figure 2: Mean of error measures on $\boldsymbol{A}$ (left) and $\boldsymbol{B}$ (left) of the classical gradient descent method (GD) and its on-line counterpart (SGD) as functions of the number of frames T. Means are computed over 1000 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $\nu = 1$. The data size is $p = 12$ with $a = 4$ and $b = 3$ and the number of samples is $n = 8$.

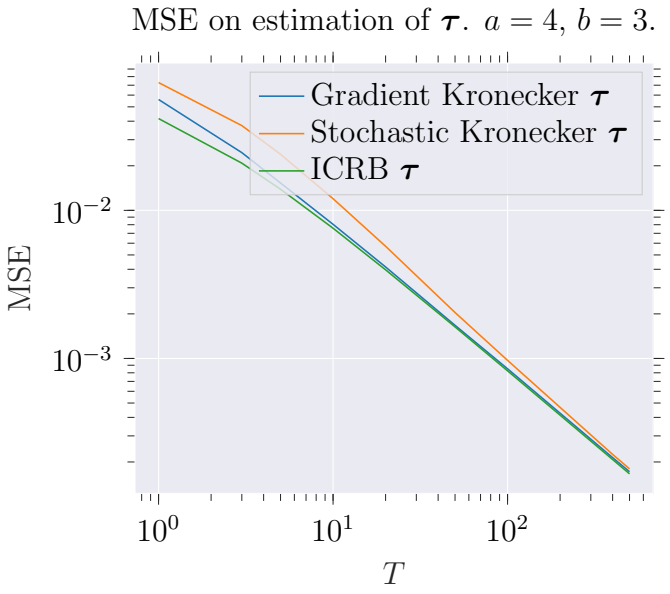


Figure 3: Mean of error measures on $\boldsymbol{\tau}$ of the classical gradient descent method (GD) and its on-line counterpart (SGD) as functions of the number of frames T. Means are computed over 1000 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $\nu = 1$. The data size is $p = 12$ with $a = 4$ and $b = 3$ and the number of samples is $n = 8$.

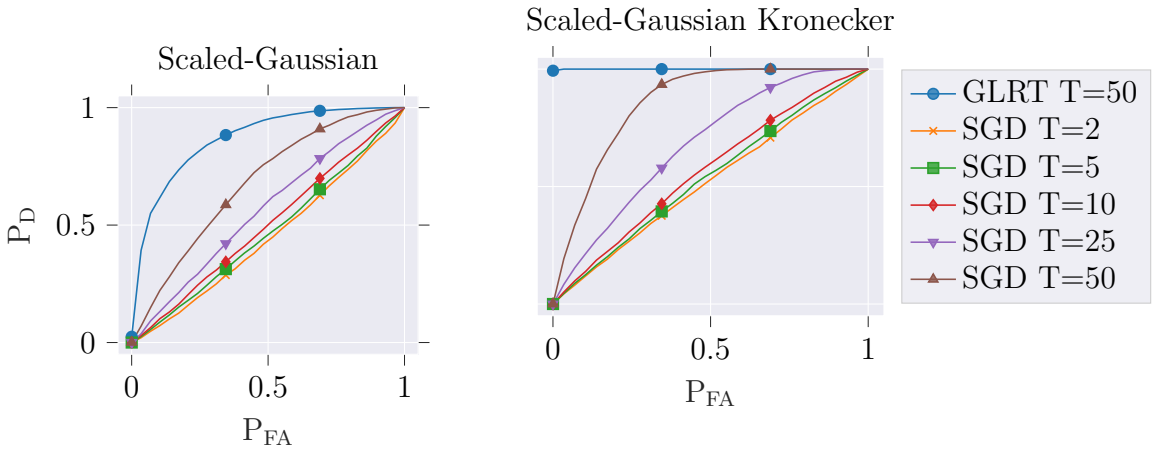


Figure 4: Gaussian data with $a = 3$, $b = 4$, $n = 13$. For **A**: $\rho_0 = 0.3+0.7j$, $\rho_1 = 0.3+0.5j$. For **B**: $\rho_0 = 0.3 + 0.6j$, $\rho_1 = 0.4 + 0.5j$. 5000 trials.

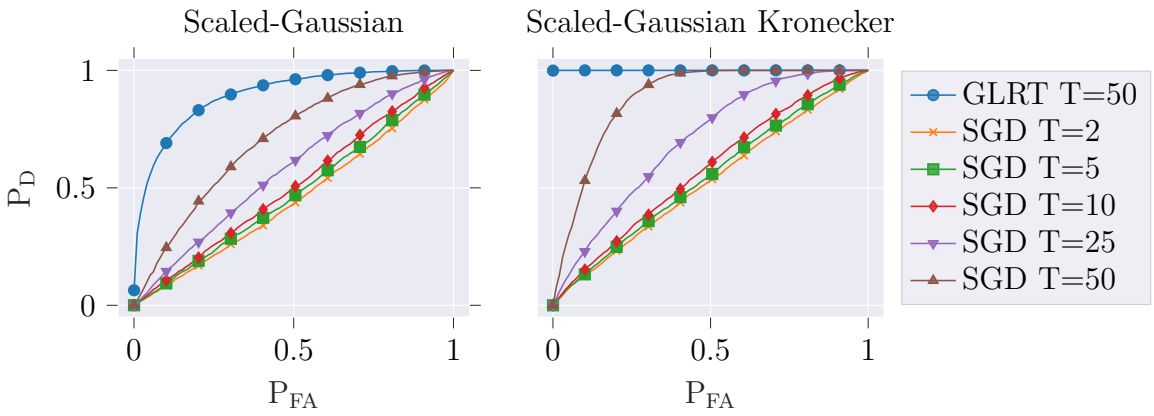


Figure 5: Non-Gaussian data with $a = 3$, $b = 4$, $n = 13$, $\nu = 100$. For **A**: $\rho_0 = 0.3 + 0.7j$, $\rho_1 = 0.3 + 0.5j$. For **B**: $\rho_0 = 0.3 + 0.6j$, $\rho_1 = 0.4 + 0.5j$. 5000 trials.

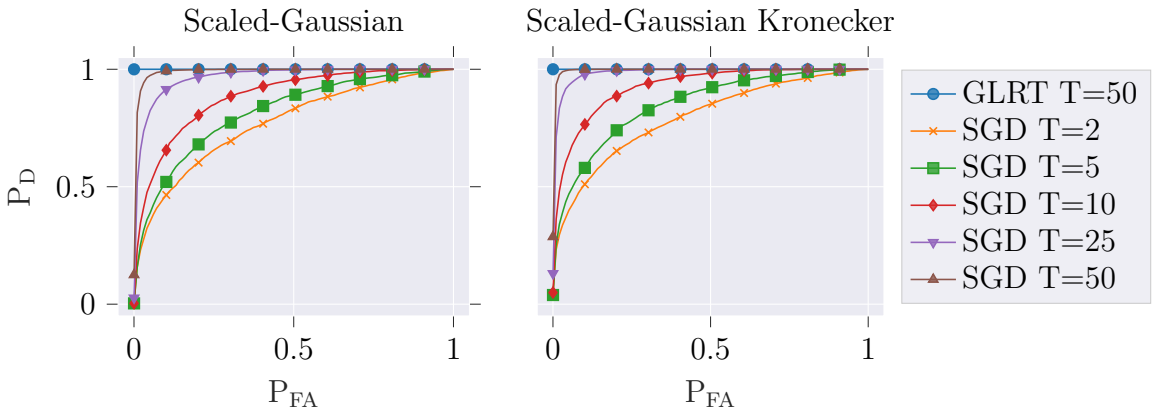
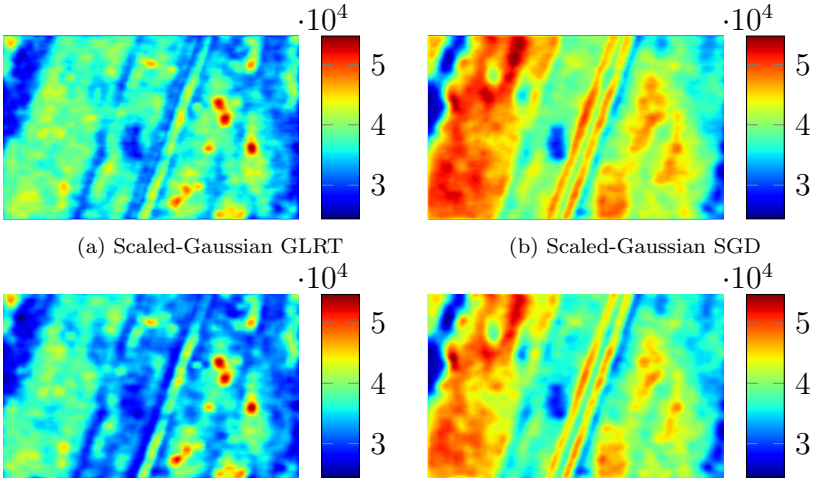


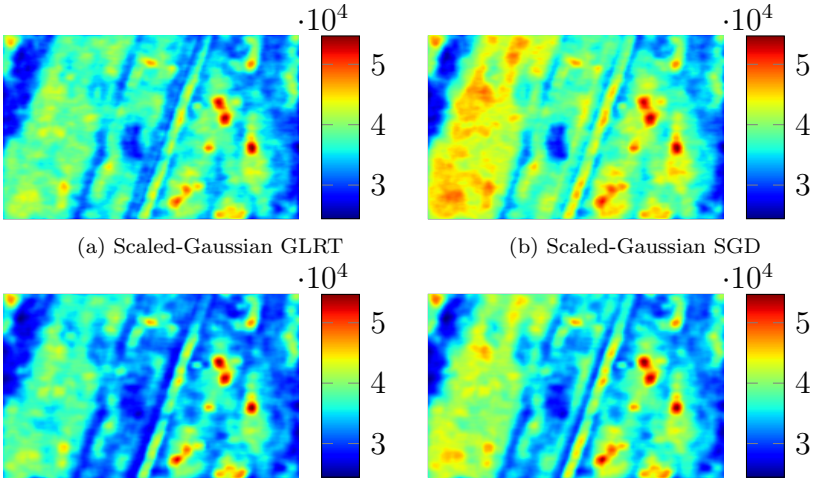
Figure 6: Non-Gaussian data with $a = 3$, $b = 4$, $n = 13$, $\nu = 1$. For **A**: $\rho_0 = 0.3 + 0.7j$, $\rho_1 = 0.3 + 0.5j$. For **B**: $\rho_0 = 0.3 + 0.6j$, $\rho_1 = 0.4 + 0.5j$. 5000 trials.



(a) Scaled-Gaussian GLRT (b) Scaled-Gaussian SGD

(c) Scaled-Gaussian Kronecker GLRT (d) Scaled-Gaussian Kronecker SGD

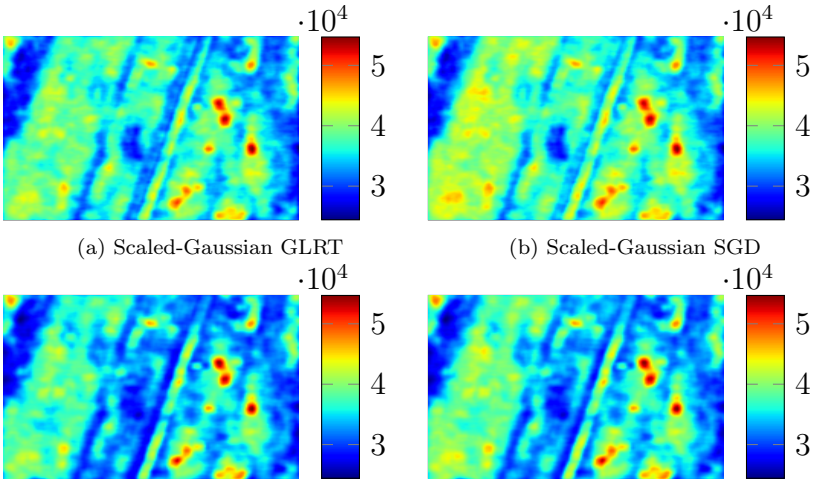
Figure 7: Window size of 7×7. No repetitions of time series: $T = 68$.



(a) Scaled-Gaussian GLRT (b) Scaled-Gaussian SGD

(c) Scaled-Gaussian Kronecker GLRT (d) Scaled-Gaussian Kronecker SGD

Figure 8: Window size of 7×7. 5 repetitions of time series: $T = 340$.



(a) Scaled-Gaussian GLRT (b) Scaled-Gaussian SGD

(c) Scaled-Gaussian Kronecker GLRT (d) Scaled-Gaussian Kronecker SGD

Figure 9: Window size of 7×7. 10 repetitions of time series: $T = 680$.