Online change point detection on Riemannian manifolds with karcher mean estimates

Xiuheng Wang, Ricardo Augusto Borsoi, Cédric Richard

ONLINE CHANGE POINT DETECTION ON RIEMANNIAN MANIFOLDS WITH KARCHER MEAN ESTIMATES

Xiuheng Wang [*], *Ricardo Augusto Borsoi* [†], *Cédric Richard* [*]

[*] Université Côte d'Azur, CNRS, OCA, France

[†] Université de Lorraine, CNRS, CRAN, Vandoeuvre-lès-Nancy, France

xiuheng.wang@oca.eu, ricardo.borsoi@univ-lorraine.fr, cedric.richard@unice.fr

ABSTRACT

Online detection of abrupt changes in streaming time series is a challenging problem with many applications, in particular when little prior knowledge of the statistics of the data is available and computation resources are scarce. While many algorithms have been developed for Euclidean spaces, there is a wealth of data that belongs to Riemannian manifolds. Taking the geometry of the data space into account is however paramount in designing effective change point detection algorithms. In this paper, we propose a non-parametric online algorithm to detect abrupt changes in manifold-valued data streams. The proposed method monitors abrupt changes in the Karcher mean of the data using a stochastic Riemannian optimization algorithm. Experiments with both synthetic and real data illustrate the performance of the proposed method.

Index Terms— Online change point detection, non-parametric, Riemannian manifolds, Karcher mean.

1. INTRODUCTION

Change-point detection (CPD) aims to detect abrupt changes in the data distribution, and is recognized as one of the most significant tasks in time series data analysis. Despite the huge literature on offline CPD, online CPD still suffers from major challenges while it plays a fundamental role in a wide range of applications such as audio [1] and video [2] segmentation, medical condition monitoring [3], or human behavior analysis [4] to cite a few. Considering whether prior knowledge about the data distributions is available or not, online CPD approaches can be divided into parametric and non-parametric. Examples of parametric methods include the cumulative sum (CUSUM) [5] and the generalized likelihood ratio test (GLRT) [6]. Such methods assume that the distribution of the data belongs to a known parametric family. However, knowledge of the data distribution is not always available, making the use of non-parametric methods necessary. Classic non-parametric strategies include monitoring changes in the mean or the variance (e.g., the exponentially weighted moving average – EWMA) [7] or in a generalized statistic [8] of the data stream. In [9], a new algorithm called NEWMA was proposed. It consists of comparing two EWMA of the data statistics computed using two distinct forgetting factors to detect change points without requiring the storage of old samples. A non-parametric online algorithm was designed in [10] based on an adaptive kernel-based density ratio estimation. Recently, deep learning has been also considered in non-parametric online CPD [11, 12].

Modern signal processing tasks increasingly tackle data that does not reside in Euclidean spaces, such as graphs or categorical data. In particular, Riemannian manifolds [13, 14] have drawn significant attention due to their widespread applications, including diffusion tensor imaging [15] and pedestrian detection [16]. However, developing methods that can process manifold-valued data is still challenging since one has to account for the nonlinear geometry of the space. Moreover, manifolds lack a vector space structure, which makes it hard to generalize algorithms originally developed for Euclidean spaces. Aside from some online CPD algorithms that have been extended to particular non-Euclidean domains such as graphs [17, 18] or categorical data [19], few works have investigated manifold-valued data. For instance, an online CPD algorithm in [20] was specifically designed for the compound Gaussian distribution. This technique is however parametric and not broadly applicable. An example of non-parametric techniques can be found in [21], but it is only able to detect a single change point in an offline manner.

In this work, we introduce a unified framework for online CPD on Riemannian manifolds based on Karcher mean estimation. Specifically, a non-parametric strategy is considered by monitoring the Karcher mean of manifold-valued data, which is estimated efficiently in an online way using a Riemannian stochastic gradient descent (SGD) algorithm. To detect abrupt change points, two Karcher mean estimates with different step sizes – one which takes longer to converge and focus on a long-term trend, and one which converges faster to assimilate change points quickly – are compared to form a test statistic. We then illustrate the proposed framework to detect changes on the Riemannian manifold of symmetric positive definite (SPD) matrices. Experimental results on sequences

The work of C. Richard was funded in part by the ANR under grant ANR-19-CE48-0002, and by the 3IA Côte d'Azur Senior Chair program.

of SPD matrices generated both synthetically and as feature descriptors of video sequences demonstrate its effectiveness.

2. BACKGROUND

This section introduces some basics of Riemannian geometry with geometrical tools for optimization. Extensive presentations can be found in [13, 14]. Then this section focuses on a usual example of Riemannian manifolds, that of $p \times p$ SPD matrices, denoted as $\mathcal{S}_p^{++}$, with an invariant metric.

2.1. Riemannian geometry and optimization

A *Riemannian manifold* $(\mathcal{M}, g)$ is defined by a constrained set $\mathcal{M}$ equipped with a *Riemannian metric* $g_x(\cdot,\cdot)$, that is, a smoothly varying inner product $\langle \cdot,\cdot \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \mapsto \mathbb{R}$, defined for each $x \in \mathcal{M}$, where $T_x\mathcal{M}$ is called the *tangent space* of $\mathcal{M}$ at x. The length of a parameterized curve, say, $c : [a, b] \mapsto \mathcal{M}$, is given by:

$$L(c) = \int_a^b \sqrt{g(\dot{c}(\alpha), \dot{c}(\alpha))} d\alpha = \int_a^b \|\dot{c}(\alpha)\| d\alpha, \quad (1)$$

where $\dot{c} = \partial c/\partial\alpha$ is the *velocity* of c. This allows us to define the *geodesic* $\gamma : [0, 1] \mapsto \mathcal{M}$, which is the unique curve of minimal length linking x and y, with $x = \gamma(0)$ and $y = \gamma(1)$. The *Riemannian distance* $d_{\mathcal{M}}(\cdot,\cdot) : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}$ is defined as follows:

$$d_{\mathcal{M}}(x, y) = \int_0^1 \|\dot{\gamma}(\alpha)\| d\alpha. \quad (2)$$

Note that $d_{\mathcal{M}}$ satisfies all conditions to be a metric.

The *exponential map* $w = exp_x(v)$ defines the point w of $\mathcal{M}$ located on the unique geodesic $\gamma_v(t)$ such that $\gamma_v(0) = x$, $\gamma_v'(0) = v$ and $\gamma_v(1) = w$. The inverse of the exponential map is defined as $v = \exp_x^{-1}(w)$. Since the exponential map may be hard to compute, one often resorts to a second-order approximation, called *retraction* mapping $R_x : T_x\mathcal{M} \mapsto \mathcal{M}$ at $x \in \mathcal{M}$, and which satisfies $d_{\mathcal{M}}(R_x(tv), \exp_x(tv)) = O(t^3)$.

Consider $f : \mathcal{M} \mapsto \mathbb{R}$ a smooth function. The *Riemannian gradient* of f at $x \in \mathcal{M}$ is the defined as the unique tangent vector $\nabla f(x) \in T_x\mathcal{M}$ satisfying:

$$\frac{d}{dt}\Big|_{t=0} f(\exp_x(tv)) = \langle \nabla f(x), v \rangle_x \quad (3)$$

for all $v \in T_x\mathcal{M}$.

2.2. The Riemannian manifold of SPD matrices

The geodesic distance of $\mathcal{S}_p^{++}$ between two SPD matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2 \in \mathcal{S}_p^{++}$ can be computed in closed form [15] as:

$$d_{\mathcal{S}_p^{++}}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \left\| \log(\mathbf{\Sigma}_2^{-\frac{1}{2}} \mathbf{\Sigma}_1 \mathbf{\Sigma}_2^{-\frac{1}{2}}) \right\|_F, \quad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

The Riemannian gradient ∇f at $\mathbf{\Sigma} \in \mathcal{S}_p^{++}$ is given by:

$$\nabla f(\mathbf{\Sigma}) = \mathbf{\Sigma}\, \mathrm{sym}(\boldsymbol{G})\mathbf{\Sigma}, \quad (5)$$

with $\boldsymbol{G} \in \mathbb{R}^{p\times p}$ the Euclidean gradient of function f at $\mathbf{\Sigma}$ and $\mathrm{sym}(\boldsymbol{G}) = \frac{1}{2}(\boldsymbol{G}^T + \boldsymbol{G})$. In practice, the Euclidean gradient can be easily computed using automatic differentiation tools. Let $\boldsymbol{\xi} \in T_{\mathbf{\Sigma}}\mathcal{S}_p^{++}$. A retraction $R_{\mathbf{\Sigma},\mathcal{S}_p^{++}} : T_{\mathbf{\Sigma}}\mathcal{S}_p^{++} \mapsto \mathcal{S}_p^{++}$ is defined as

$$R_{\mathbf{\Sigma},\mathcal{S}_p^{++}}(\boldsymbol{\xi}) = \mathbf{\Sigma} + \boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}\mathbf{\Sigma}^{-1}\boldsymbol{\xi}. \quad (6)$$

This retraction is a second-order approximation of the exponential mapping, that is,

$$\exp_{\mathbf{\Sigma},\mathcal{S}_p^{++}}(t\boldsymbol{\xi}) = R_{\mathbf{\Sigma},\mathcal{S}_p^{++}}(t\boldsymbol{\xi}) + O(t^3). \quad (7)$$

3. METHODOLOGY

Let us consider a time series of independent random variables $\{\boldsymbol{x}_t\}_{t\in\mathbb{N}}$ lying on a Riemannian manifold $\mathcal{M}$. We assume that there exists a time index $t_r \in \mathbb{N}$ with an abrupt change in the probability distribution of $\boldsymbol{x}_t$, that is,

$$t < t_r : \ \boldsymbol{x}_t \sim P_1(\boldsymbol{x}), \qquad t \geq t_r : \ \boldsymbol{x}_t \sim P_2(\boldsymbol{x}), \quad (8)$$

where $P_1(\boldsymbol{x})$ and $P_2(\boldsymbol{x})$ denote two different probability measures on $\mathcal{M}$ [22] which represent the distribution of data $\boldsymbol{x}_t$ before and after the change point t_r. Note that, to simplify the presentation, (8) considers only a single change point. However, the algorithm presented hereafter can handle multiple change points.

CPD algorithms aim to estimate a change point $\hat{t}_r$ according to two complementary objectives: 1) minimizing the detection delay, i.e., $\hat{t}_r - t_r$ for $\hat{t}_r$ being the first detection after t_r; and 2) minimizing the probability of *false alarms*, i.e., of flagging some $t \neq t_r$ as a change point. In this work, we consider a problem setting where manifold-valued data $\boldsymbol{x}_t$ are observed sequentially over time and change points must be detected *online*. This means that we need to decide if each time instant $t \in \mathbb{N}$ is a change point based only on past data $\{\boldsymbol{x}_{t'}\}_{t'\leq t}$. Moreover, unlike [20], we focus on non-parametric strategies, which do not make additional assumptions about the statistical distribution of the data.

3.1. Non-parametric statistics and the Karcher mean

As discussed above, we focus on *non-parametric* strategies, where there is no prior knowledge about probability measures of the data stream. In Euclidean spaces, this can be done by monitoring changes in the mean or the variance [7], or in a generalized statistic [8] of the data stream. In order to generalize such strategies to Riemannian manifolds, we consider

the Karcher mean [23] of the data stream $\boldsymbol{x}_t \in \mathcal{M}$, which provides a generalization of the center of mass from Euclidean domains to a manifold $\mathcal{M}$.

The Karcher mean is a generalization of the Fréchet mean, which is defined as the set of values that (globally) minimize the expected variance:

$$f(\boldsymbol{m}) = \mathbb{E}_{\boldsymbol{x}\sim P(\boldsymbol{x})}\big\{d^2_{\mathcal{M}}(\boldsymbol{m},\boldsymbol{x})\big\} = \int d^2_{\mathcal{M}}(\boldsymbol{m},\boldsymbol{x}) dP(\boldsymbol{x})$$

of the Riemannian distance $d_{\mathcal{M}}$, that is,

$$\boldsymbol{m}^* \in \arg\min_{\boldsymbol{m}} \; f(\boldsymbol{m})\,. \tag{9}$$

Note that the existence and uniqueness of the Fréchet mean are not guaranteed. The Karcher mean relaxes this definition by considering the local optima of $f(\boldsymbol{m})$, instead of only the global one. This allows to establish existence and uniqueness conditions [24], and also makes it possible to compute $\boldsymbol{m}$ by locally solving (9) using Riemannian optimization methods [22]. The Karcher mean is unique in many manifolds, such as those that are connected and have nonpositive curvature [25], which includes $\mathcal{S}_p^{++}$.

The CPD strategy on manifolds proposed in the following monitors abrupt changes in the Karcher mean of the data stream. An important requirement is that change points must be detected in an *online* way, that is, only based on past data. Consequently, we start the presentation with an online Riemannian descent algorithm to estimate the Karcher mean of streaming data. This will be an integral part of the CPD strategy presented afterward.

3.2. Online estimation of the Karcher mean

In optimization problem (9), the cost function cannot be computed explicitly because $P(\boldsymbol{x})$ is unknown. However, observations $\{\boldsymbol{x}_t\}$ are available to compute function $d^2_{\mathcal{M}}(\boldsymbol{m},\boldsymbol{x}_t)$ for any parameter $\boldsymbol{m}$ and data point $\boldsymbol{x}_t$. That function can be viewed as a stochastic approximation of the loss $f(\boldsymbol{m})$ updated with new input data $\boldsymbol{x}_t$. Consequently, we consider using the Riemannian SGD algorithm [26] to address problem (9). On manifold $\mathcal{M}$, the update of $\boldsymbol{m}$ with a step size α is given by:

$$\boldsymbol{m}_{t+1} = \exp_{\boldsymbol{m}_t}\big(-\alpha H(\boldsymbol{m}_t,\boldsymbol{x}_t)\big), \tag{10}$$

where $\exp_{\boldsymbol{m}}$ is the exponential map at $\boldsymbol{m}$, and $H(\boldsymbol{m},\boldsymbol{x})$ denotes the Riemannian gradient of the loss such that

$$\mathbb{E}_{\boldsymbol{x}\sim P(\boldsymbol{x})}\big\{H(\boldsymbol{m},\boldsymbol{x})\big\} = \int H(\boldsymbol{m},\boldsymbol{x}) dP(\boldsymbol{x}) = \nabla f(\boldsymbol{m}).$$

For computational simplicity, we replace the exponential map in (10) by a retraction $R_{\boldsymbol{m}_t}$. This yields the alternative update:

$$\boldsymbol{m}_{t+1} = R_{\boldsymbol{m}_t}\big(-\alpha H(\boldsymbol{m}_t,\boldsymbol{x}_t)\big). \tag{11}$$

Algorithm 1: Online CPD on manifolds

Input: $\{\boldsymbol{x}_t\}$, step sizes λ, Λ, threshold ξ.
Initialization: $\boldsymbol{m}_{\lambda,0} = \boldsymbol{m}_{\Lambda,0} = \boldsymbol{x}_0$;
for $t = 1, 2, 3, \ldots$ **do**
  Update the "fast" and "slow" Karcher mean estimates $\boldsymbol{m}_{\lambda,t}$ and $\boldsymbol{m}_{\Lambda,t}$ using (12) and (13) ;
  Compute the test statistic $g_t = d_{\mathcal{M}}(\boldsymbol{m}_{\lambda,t}, \boldsymbol{m}_{\Lambda,t})$;
  if $g_t > \xi$ **then**
    Flag t as a change point;
  end
end

3.3. An adaptive CPD with the Karcher mean

We aim to detect change points by monitoring abrupt changes in $\boldsymbol{m}$ over time, that is, a point t' is labeled as a change point if $\boldsymbol{m}$ changed abruptly at t'. This requires knowledge of two quantities of interest, $\boldsymbol{m}_{\text{bef}}$ and $\boldsymbol{m}_{\text{aft}}$, which correspond to the Karcher mean before and after a candidate change point t'. First, we propose to compute estimates of these values, say $\widehat{\boldsymbol{m}}_{\text{bef}}$ and $\widehat{\boldsymbol{m}}_{\text{aft}}$. Then, as a test statistic, we propose to compare these two quantities using the Riemannian distance, that is, $d_{\mathcal{M}}(\widehat{\boldsymbol{m}}_{\text{bef}}, \widehat{\boldsymbol{m}}_{\text{aft}})$: the larger the Riemannian distance between the Karcher mean estimates before and after instant t', the more likely we are to flag t' as a change point.

The question is how to calculate $\widehat{\boldsymbol{m}}_{\text{bef}}$ and $\widehat{\boldsymbol{m}}_{\text{aft}}$ efficiently and in an online way. Previous work proposed to partition a data stream of length N into two segments, $\{1,\ldots,t'-1\}$ and $\{t',\ldots,N\}$ for every t', and testing for differences between their Karcher mean and variance [21]. However, this strategy cannot process data streams on-the-fly or detect multiple change points. In [9], within the realm of Euclidean geometry, these estimates were computed using two exponentially weighted moving averages with different forgetting factors. Nevertheless, this principle cannot be transposed directly to Riemannian manifold setting. Instead, we propose to use two estimates provided by Riemannian stochastic gradient descent algorithms as presented in Section 3.2, with two different fixed step sizes $\lambda < \Lambda$. The Karcher means are updated according to (11) as follows:

$$\boldsymbol{m}_{\lambda,t+1} = R_{\boldsymbol{m}_{\lambda,t}}\big(-\lambda H(\boldsymbol{m}_{\lambda,t},\boldsymbol{x}_t)\big)\,, \tag{12}$$

$$\boldsymbol{m}_{\Lambda,t+1} = R_{\boldsymbol{m}_{\Lambda,t}}\big(-\Lambda H(\boldsymbol{m}_{\Lambda,t},\boldsymbol{x}_t)\big)\,, \tag{13}$$

with initialization $\boldsymbol{m}_{\lambda,0} = \boldsymbol{m}_{\Lambda,0} = \boldsymbol{x}_0$. Convergence of the updates (12) and (13) is directly affected by λ and Λ. Constraint $\lambda < \Lambda$ thus means that $\boldsymbol{m}_{\Lambda,t}$ is more likely to adapt to new data and approximates $\widehat{\boldsymbol{m}}_{\text{aft}}$, while $\boldsymbol{m}_{\lambda,t}$ has longer memory and is more suitable to estimate baseline trend $\widehat{\boldsymbol{m}}_{\text{bef}}$.

Using these two estimates, we can define an adaptive CPD statistic by comparing the difference between $\boldsymbol{m}_{\lambda,t}$ and $\boldsymbol{m}_{\Lambda,t}$ using the Riemannian distance on $\mathcal{M}$ as:

$$g_t = d_{\mathcal{M}}(\boldsymbol{m}_{\lambda,t}, \boldsymbol{m}_{\Lambda,t})\,. \tag{14}$$

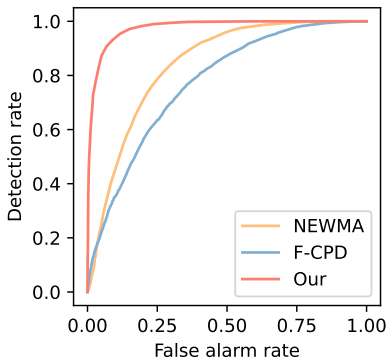

Fig. 1. ROC curves for all compared algorithms.

CPD is then performed by comparing g_t to a threshold ξ. The full CPD procedure is summarized in Algorithm 1.

4. EXPERIMENTS

We shall now present some experiments on manifold $\mathcal{S}_p^{++}$ discussed in Section 2. We consider the problem of detecting change points with Algorithm 1 both in a sequence of synthetic SPD matrices, and in a sequence of region covariance descriptors derived from a real video. With $\{\boldsymbol{\Sigma}_t\}_{t\in\mathbb{N}}$ lying on $\mathcal{S}_p^{++}$ and the metric defined in (4), the Karcher means were estimated by minimizing the objective function $f(\boldsymbol{\Sigma}) = \mathbb{E}_{\boldsymbol{S}\sim P(\boldsymbol{S})}\{\left\|\log(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{S}^{-\frac{1}{2}})\right\|_F^2\}$ using the Riemannian SGD algorithms in (12) and (13) and the stochastic approximation $\boldsymbol{S} \simeq \boldsymbol{\Sigma}_t$ at each t. They were used to compute the online CPD statistic in (14).

We compared our method with two baseline algorithms, namely, NEWMA [9] and the Fréchet CPD (F-CPD) [21]. On the one hand, since NEWMA was originally designed for Euclidean spaces, we applied it to the vectorization of the left triangular and diagonal parts of each SPD matrix. On the other hand, F-CPD was designed for manifold-valued data but can only detect a single change point in an offline manner. We tackled these issues by using F-CPD over consecutive sliding windows of length 100. We set $\lambda = 0.01$ and $\Lambda = 0.02$ for synthetic data, and $\lambda = 0.05$ and $\Lambda = 0.06$ for real data.

Experiment with synthetic data: The synthetic matrices $\boldsymbol{\Sigma}_t \in \mathcal{S}_p^{++}$ with $p = 6$ were sampled from a Wishart distribution with the scaling matrix $\boldsymbol{V}$ and d degrees of freedom. We generated 800 samples and set a change point at $t_r = 500$ where we reset $\boldsymbol{V}$. Fig. 1 shows the Receiver Operating Characteristic (ROC) curves of all methods for 5000 Monte Carlo runs. It can be seen that our method achieved a significant improvement in the detection rate with a low rate of false alarms when compared to both NEWMA, which does not take the manifold geometry into account, and to F-CPD, which was designed to operate offline.

Experiment with real data: To further evaluate our approach, we made use of a real video of an outdoor scene

Fig. 2. Snapshot of the video sequence with its super-pixels decomposed via the SLIC algorithm [27].

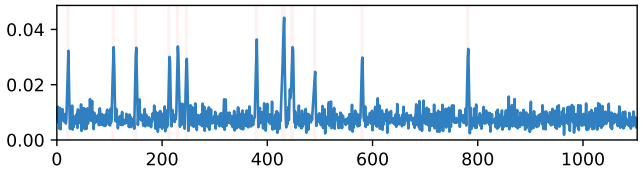

Fig. 3. Test statistic applied to a super-pixel of the video sequence. The ground-truth change points are colored in red.

from [28]. This scene contains intermittent object motions including cars and pedestrians as well as subtle, non-informative changes such as tree leaves moving. The video contains color images with 658×491 pixels. Ground truths of segmentation results for larger moving targets are available. With ground truths, the first $T = 1150$ sequential frames were selected and cropped to 658×260 pixels. We considered detecting change points in disjoint, compact regions of this scene by segmenting each frame into superpixels via the SLIC technique [27]. A snapshot of this scene with its superpixel decomposition is shown in Fig. 2.

We considered the region covariance descriptor features in [16] but with the pixel location information removed. Let $\{\boldsymbol{s}_t\}_{t\in\mathbb{N}}$ denote the video sequence. Each frame $\boldsymbol{s}_t$ was first processed by computing a feature vector for each pixel:

$$\boldsymbol{z}_t(x,y) = \left[|I_x|, |I_y|, \sqrt{I_x^2 + I_y^2}, |I_{xx}|, |I_{yy}|, \arctan\left(\frac{I_x}{I_y}\right)\right]^T$$

where (x, y) represents pixel locations and I_x, I_y, I_{xx}, I_{yy} denote intensity derivatives. For each superpixel $\boldsymbol{p}$ containing n pixels, the covariance descriptor was an SPD matrix of the feature vectors $\boldsymbol{z}_t(x, y)$ computed as follows:

$$\boldsymbol{\Sigma}_t = \frac{1}{n-1}\sum_{(x,y)\in\boldsymbol{p}} \big(\boldsymbol{z}_t(x,y) - \bar{\boldsymbol{z}}_t\big)\big(\boldsymbol{z}_t(x,y) - \bar{\boldsymbol{z}}_t\big)^T,$$

with $\boldsymbol{\Sigma}_t \in \mathcal{S}_p^{++}$ and $p = 6$, and $\bar{\boldsymbol{z}}_t$ the sample mean of $\boldsymbol{z}_t(x, y)$ for all $(x, y) \in \boldsymbol{p}$.

We used Algorithm 1 to detect change points in the data stream $\{\boldsymbol{\Sigma}_t\}$. The resulting test statistic for one given superpixel and the corresponding change points are illustrated in Figure 3, and compared to the ground truth. We can observe that all peaks of the test statistic are located near change

points pointed out by the ground truth, which indicates that the proposed method can reach a low false alarm rate.

5. CONCLUSION

In this paper, we presented a general approach for the online detection of change points in Riemannian manifolds based on Karcher mean estimation. An adaptive test statistic was computed by comparing two Karcher means estimated with Riemannian SGD algorithms, one converging faster to assimilate new data, and another one converging more slowly to focus on a long-term trend. Experimental results on the Riemannian manifold of SPD matrices illustrated the superiority of our strategy compared with two baseline algorithms that either operate offline or do not take manifold geometry into account.

6. REFERENCES

[1] A. Bietti, F. Bach, and A. Cont, "An online EM algorithm in hidden (semi-) Markov models for audio segmentation and clustering," in *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2015, pp. 1881–1885.

[2] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, 2014.

[3] R. Malladi, G. P. Kalamangalam, and B. Aazhang, "Online Bayesian change point detection algorithms for segmentation of epileptic activity," in *Proc. Asilomar conference on signals, systems and computers*. IEEE, 2013, pp. 1833–1837.

[4] S. Aminikhanghahi, T. Wang, and D. J. Cook, "Real-time change point detection with application to smart home time series data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 1010–1023, 2018.

[5] C. Inclan and G. C. Tiao, "Use of cumulative sums of squares for retrospective detection of changes of variance," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 913–923, 1994.

[6] F. Gustafsson, "The marginalized likelihood ratio test for detecting abrupt changes," *IEEE Transactions on automatic control*, vol. 41, no. 1, pp. 66–78, 1996.

[7] A. Costa and M. Rahim, "A single ewma chart for monitoring process mean and process variance," *Quality Technology & Quantitative Management*, vol. 3, no. 3, pp. 295–305, 2006.

[8] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample-problem," *Advances in neural information processing systems*, vol. 19, 2006.

[9] N. Keriven, D. Garreau, and I. Poli, "NEWMA: a new method for scalable model-free online change-point detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3515–3528, 2020.

[10] A. Ferrari, C. Richard, A. Bourrier, and I. Bouchikhi, "Online change-point detection with kernels," *Pattern Recognition*, p. 109022, 2022.

[11] Z. Atashgahi, D. C. Mocanu, R. Veldhuis, and M. Pechenizkiy, "Memory-free online change-point detection: A novel neural network approach," *arXiv preprint arXiv:2207.03932*, 2022.

[12] X. Wang, R. A. Borsoi, C. Richard, and J. Chen, "Change point detection with neural online density-ratio estimator," in *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2023.

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

[14] N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, 2023.

[15] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of computer vision*, vol. 66, no. 1, pp. 41–66, 2006.

[16] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[17] A. Ferrari and C. Richard, "Non-parametric community change-points detection in streaming graph signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5545–5549.

[18] R. A. Borsoi, C. Richard, A. Ferrari, J. Chen, and J. C. M. Bermudez, "Online graph-based change point detection in multiband image sequences," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 850–854.

[19] D. Ienco, A. Bifet, B. Pfahringer, and P. Poncelet, "Change detection in categorical evolving data streams," in *29th annual ACM symposium on applied computing*, 2014, pp. 792–797.

[20] F. Bouchard, A. Mian, J. Zhou, S. Said, G. Ginolhac, and Y. Berthoumieu, "Riemannian geometry for compound gaussian distributions: Application to recursive change detection," *Signal Processing*, vol. 176, pp. 107716, 2020.

[21] P. Dubey and H.-G. Müller, "Fréchet change-point detection," *The Annals of Statistics*, vol. 48, no. 6, pp. 3312–3335, 2020.

[22] X. Pennec, "Probabilities and statistics on riemannian manifolds: A geometric approach," Tech. Rep. 5093, INRIA, 2004.

[23] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[24] W. S. Kendall, "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence," *Proceedings of the London Mathematical Society*, vol. 3, no. 2, pp. 371–406, 1990.

[25] B. Afsari, "Riemannian ℓ^p center of mass: existence, uniqueness, and convexity," *Proceedings of the American Mathematical Society*, vol. 139, no. 2, pp. 655–673, 2011.

[26] S. Bonnabel, "Stochastic gradient descent on riemannian manifolds," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013.

[27] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[28] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *IEEE computer society conference on computer vision and pattern recognition workshops*, 2012, pp. 1–8.