

RIEMANNIAN CHANGE POINT DETECTION ON MANIFOLDS WITH ROBUST CENTROID ESTIMATION

Xiuheng Wang^{}, Ricardo Borsoi^{*}, Arnaud Breloy[†], Cédric Richard[‡]*

^{*} Université de Lorraine, CNRS, CRAN, France

[†] Conservatoire National des Arts et Métiers, CEDRIC, France

[‡] Université Côte d’Azur, CNRS, OCA, France

dr.xiuheng.wang@gmail.com, ricardo.borsoi@univ-lorraine.fr, arnaud.breloy@lecnam.net, cedric.richard@unice.fr

ABSTRACT

Non-parametric change-point detection in streaming time series data is a long-standing challenge in signal processing. Recent advancements in statistics and machine learning have increasingly addressed this problem for data residing on Riemannian manifolds. One prominent strategy involves monitoring abrupt changes in the center of mass of the time series. Implemented in a streaming fashion, this strategy, however, requires careful step size tuning when computing the updates of the center of mass. In this paper, we propose to leverage robust centroid on manifolds from M-estimation theory to address this issue. Our proposal consists of comparing two centroid estimates: the classical Karcher mean (sensitive to change) versus one defined from Huber’s function (robust to change). This comparison leads to the definition of a test statistic whose performance is less sensitive to the underlying estimation method. We propose a stochastic Riemannian optimization algorithm to estimate both robust centroids efficiently. Experiments conducted on both simulated and real-world data across two representative manifolds demonstrate the superior performance of our proposed method.

Index Terms— Non-parametric, change point detection, Riemannian manifolds, robust centroid estimation, M-estimator, online, stochastic optimization.

1. INTRODUCTION

Parametric and non-parametric CPD: The change-point detection (CPD) problem aims to identify abrupt changes in the statistical distribution of time series data. CPD has an extensive history within the statistics literature, with diverse applications ranging from medical [1] and speech processing [2]

to image analysis [3]. Various CPD algorithms have been proposed depending on the level of prior information available. Classical CPD approaches typically utilize parametric strategies, requiring knowledge of the specific form of the probability density function (PDF) before and after the change point. Notable parametric CPD algorithms include the cumulative sum (CUSUM) [4], designed for detecting shifts in either the mean or variance, and the generalized likelihood ratio test (GLRT) [5], which models the data PDF using a linear state-space representation.

While strategies based on parametric statistical models can offer strong performance guarantees, non-parametric settings have been the subject of increasing interest since they do not make strict assumptions on the data distribution that could make the algorithm sensitive to modeling errors [6]. Notable approaches include the Exponentially Weighted Moving Average (EWMA) [7] and kernel Maximum Mean Discrepancy (MMD) [8]. Recently, the NEWMA algorithm [9] was introduced. It detects change points by comparing two EWMA streaming statistics [7], each computed with different forgetting factors. The kernel MMD statistic, originally introduced for hypothesis testing in [8], has also found extensive application in kernel CPD [10]. Another kernel-based algorithm, which makes use of adaptive density ratio estimation, was developed in [11]. Finally, the potential of neural networks was investigated in [12].

CPD on manifolds: The extension of nonparametric CPD algorithms to handle data defined on non-Euclidean spaces, such as Riemannian manifolds, has been the subject of recent research [13–15]. Formally, given a time series of independent random variables $\{\mathbf{x}_t\}_{t \in \mathbb{N}}$ lying on a Riemannian manifold \mathcal{M} , the Riemannian CPD problem [14, 15] consists of detecting a time index $t_r \in \mathbb{N}$ at which the distribution of the data abruptly changes:

$$t < t_r : \mathbf{x}_t \sim P_1(\mathbf{x}), \quad t \geq t_r : \mathbf{x}_t \sim P_2(\mathbf{x}), \quad (1)$$

where $P_1(\mathbf{x})$ and $P_2(\mathbf{x})$ denote the probability measure of the data $\mathbf{x}_t \in \mathcal{M}$ before and after the change point t_r , re-

The work of Cédric Richard was supported in part by the French Government through the 3IA Côte d’Azur Investments in the Future Project under grant ANR-19-P3IA-0002, and in part by grant ANR-19-CE48-0002. The work of Ricardo Borsoi was supported in part by the French National Research Agency, under grants ANR-23-CE23-0024, ANR-23-CE94-0001, and by the National Science Foundation, under grant NSF 2316420.

spectively. A generalization of problem (1) is the *multiple CPD problem*, where one aims to estimate not a single but a set of change points $\{t_{r_1}, t_{r_2}, \dots\} \subset \mathbb{N}_*$, meaning that \mathbf{x}_t is allowed to change at multiple locations.

Existing algorithms and contribution: Because they do not account for the geometry of the data residing on \mathcal{M} , particularly the absence of a vector space structure, standard nonparametric algorithms are not suitable for addressing the problem (1). For manifold-valued data, recent works have proposed to monitor abrupt changes in the Karcher mean [16] on \mathcal{M} , defined as:

$$\mathbf{m}^* = \arg \min_{\mathbf{m} \in \mathcal{M}} f(\mathbf{m}), \quad (2)$$

which is the point that minimizes the Karcher variance:

$$f(\mathbf{m}) = \mathbb{E}_{\mathbf{x} \sim P(\mathbf{x})} \left\{ \frac{1}{2} d_{\mathcal{M}}^2(\mathbf{m}, \mathbf{x}) \right\}.$$

For example, an offline technique monitoring changes in \mathbf{m}^* before and after change points was proposed in [13]. More recently, an online algorithm [14, 15] extended NEWMA to general manifolds by comparing two recursive estimates of \mathbf{m}^* computed via stochastic gradient algorithms. However, none of these methods considered a more general alternative: recursively monitoring two *different* statistics of the data \mathbf{x}_t constructed to be similar in the absence of a change point and significantly different after a change point occurs.

A key concept in our work is to leverage a robust centroid estimator on \mathcal{M} . Based on the theory of M-estimators, a class of robust centroid estimators arises from the so-called Huber function [17, 18] and covers the concept of Karcher mean. Robust centroid estimation has recently been applied to various tasks in manifolds [19, 20]. The primary contribution of this paper is the introduction of robust centroid estimation into Riemannian CPD. We specially explore two special cases of the Huber centroid, one prioritizing adaptability and the other emphasizing robustness. The comparison between these two centroids forms the basis for our test statistic, which notably reduces the dependence on the convergence of the estimation algorithm. Additionally, we propose an online estimation method for the Huber centroid using a Riemannian stochastic optimization algorithm. Our methodology is tailored explicitly for two commonly encountered manifolds: the symmetric positive definite (SPD) manifold and the Grassmann manifold. Numerical experiments on both synthetic and real-world datasets validate the effectiveness of our Riemannian CPD approach.

2. BACKGROUND

This section briefly introduces concepts of Riemannian geometry and optimization on manifolds [21, 22].

A *Riemannian manifold* (\mathcal{M}, g) is a constrained set \mathcal{M} endowed with a *Riemannian metric* $g_x(\cdot, \cdot) : T_x \mathcal{M} \times T_x \mathcal{M} \rightarrow \mathbb{R}$, defined for every point $x \in \mathcal{M}$, with $T_x \mathcal{M}$ the so-called *tangent space* of \mathcal{M} at x . A *geodesic* $\gamma_v : [0, 1] \rightarrow \mathcal{M}$ is the minimal length curve linking two points $x, y \in \mathcal{M}$ such that $x = \gamma_v(0)$ and $y = \gamma_v(1)$, with $v \in T_x \mathcal{M}$ the velocity of γ_v at 0 denoted by $\dot{\gamma}_v(0)$. The geodesic distance $d_{\mathcal{M}}(\cdot, \cdot) : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ is defined as the length of the geodesic that connects two points $x, y \in \mathcal{M}$. The *exponential map* $w = \exp_x(v)$ is defined as the point $w \in \mathcal{M}$ located on the unique geodesic $\gamma_v(t)$ with endpoints $x = \gamma_v(0)$, $w = \gamma_v(1)$ and velocity $v = \dot{\gamma}_v(0)$. Calculating the exponential map can be computationally demanding. In practice, it is common to employ a *retraction* $R_x : T_x \mathcal{M} \rightarrow \mathcal{M}$ instead, defined at every $x \in \mathcal{M}$, which consists of a second-order approximation to the exponential map, satisfying $d_{\mathcal{M}}(R_x(tv), \exp_x(tv)) = O(t^3)$. Consider a smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$. The *Riemannian gradient* of f at $x \in \mathcal{M}$ is defined as the unique tangent vector $\nabla f(x) \in T_x \mathcal{M}$ satisfying $\frac{d}{dt} \big|_{t=0} f(\exp_x(tv)) = \langle \nabla f(x), v \rangle_x$, for all $v \in T_x \mathcal{M}$.

3. METHODOLOGY

In the online setting, samples \mathbf{x}_t in the Riemannian manifold \mathcal{M} are observed sequentially over time. Thus, at each time instant $t' \in \mathbb{N}_*$ it is necessary to determine whether to flag t' as a changepoint based on previously measured data, $\{\mathbf{x}_1, \dots, \mathbf{x}_{t'}\}$. Considering the signal model (1) in which a change point occurs at time t_r , the detection problem in our setting involves two distinct objectives, the first being to minimize the mean detection delay (MDD) $\mathbb{E}\{\hat{t}_r - t_r\}$, with \hat{t}_r being the first detection after the true change point t_r , while the second is to maximize the average run time (ARL) $\mathbb{E}\{\hat{t}_r\}$, where $\hat{t}_r < t_t$ denotes the time at which the first false alarm is flagged.

3.1. Robust centroid estimation

To detect change points on \mathcal{M} , we propose monitoring the robust centroid of the data stream $\mathbf{x}_t \in \mathcal{M}$, extending the framework of the Karcher mean defined in (2). Based on the theory of M-estimation, the robust centroid can be obtained by minimizing the following cost function:

$$\mathbf{m}_\rho^* = \arg \min_{\mathbf{m}_\rho \in \mathcal{M}} f_\rho(\mathbf{m}_\rho), \quad (3)$$

where

$$f_\rho(\mathbf{m}_\rho) = \mathbb{E}_{\mathbf{x} \sim P(\mathbf{x})} \left\{ \frac{1}{2} \rho(d_{\mathcal{M}}(\mathbf{m}_\rho, \mathbf{x})) d_{\mathcal{M}}^2(\mathbf{m}_\rho, \mathbf{x}) \right\},$$

with $\rho : \mathbb{R} \rightarrow \mathbb{R}$ a function that ensures the robustness to outliers.

Depending on ρ , various robust centroid estimators can be defined [20]. For example, the Huber's estimator [17] can be considered with a specific Huber function defined as

$$\rho(a) = \min \left(1, \frac{A}{a} \right), \quad (4)$$

where $A > 0$ is a fixed parameter.

In [14, 15], the authors consider only the Karcher mean estimator defined in (2). In this work, we extend this strategy by viewing the Karcher mean estimator as a special instance of the robust centroid estimator in (3) corresponding to the choice of $A = \infty$ in (4). We further compare this adaptive centroid estimator with another special case of the robust centroid estimator obtained by setting $A \in (0, \infty)$ in (4), which provides enhanced robustness in the presence of outliers – in our context, the post-change samples.

Although other robust centroid estimators employing different forms of ρ exist, such as ℓ_1 -norm, Cauchy–Lorentz-type and Geman–McClure-type [20], we restrict our analysis to Huber-type estimators. This choice preserves geodesic convexity and smoothness of the cost function on specific manifolds, notably the SPD manifold. This allows us to leverage the powerful Riemannian stochastic gradient descent (SGD) [23, 24] algorithm, whose convergence properties strongly rely on these characteristics of the cost function.

3.2. Riemannian SGD-based estimates

As discussed above, the proposed Riemannian CPD strategy monitors abrupt changes in the robust centroid of the data stream. A key requirement is that change points must be detected *online*, relying solely on past data. In [14, 15], within the Riemannian geometry framework, Karcher mean estimates were computed using two Riemannian SGD with distinct step sizes. However, this approach can be restrictive, as these step sizes must be selected within a limited range to ensure the convergence of the Riemannian SGD (see Theorem 4.1 in [15]). To address this issue, we propose comparing two estimates derived from two distinct special cases of the robust centroid estimator, both computed using the *same* step size, as follows.

On the one hand, when choosing $A = \infty$ in (4), the robust centroid estimator in (3) reduces to the Karcher mean defined in (2). Therefore, for this case, we consider the following iterative update using Riemannian SGD [23, 24]:

$$\mathbf{m}_{t+1} = R_{\mathbf{m}_t}(-\alpha H(\mathbf{m}_t, \mathbf{x}_t)), \quad (5)$$

where α is the step size and $H(\mathbf{m}_t, \mathbf{x}_t)$ is the stochastic Riemannian gradient of $f(\mathbf{m}_t)$, approximated using the streaming data \mathbf{x}_t . This yields a baseline estimator that, according to the theory of M-estimation, is less robust to outliers and, consequently, to change points in an online setting.

On the other hand, when $A \in (0, \infty)$, the robust centroid estimator defined in (3) can be computed using Riemannian

Algorithm 1: Riemannian CPD with robust centroid estimates

Input: $\{\mathbf{x}_t\}$, step sizes α , threshold ξ .

```

1 Initialization:  $\mathbf{m}_0 = \mathbf{m}_{\rho,0} = \mathbf{x}_0$ ;
2 for  $t = 1, 2, 3, \dots$  do
3   Update the adaptive and robust centroid estimates  $\mathbf{m}_t$ 
   and  $\mathbf{m}_{\rho,t}$  using (5) and (6);
4   Compute the test statistic  $g_t = d_{\mathcal{M}}(\mathbf{m}_t, \mathbf{m}_{\rho,t})$ ;
5   if  $g_t > \xi$  then
6     Flag  $t$  as a change point;
7   end
8 end
```

SGD through the following iterative update:

$$\mathbf{m}_{\rho,t+1} = R_{\mathbf{m}_{\rho,t}}(-\alpha H_{\rho}(\mathbf{m}_{\rho,t}, \mathbf{x}_t)), \quad (6)$$

where $H_{\rho}(\mathbf{m}_{\rho,t}, \mathbf{x}_t)$ is the stochastic Riemannian gradient of $f_{\rho}(\mathbf{m}_t)$ approximated using the streaming data \mathbf{x}_t . Using the Huber function defined in (4), the relationship between $H_{\rho}(\mathbf{m}_{\rho,t}, \mathbf{x}_t)$ and $H(\mathbf{m}_{\rho,t}, \mathbf{x}_t)$ can be derived as follows:

$$H_{\rho}(\mathbf{m}_{\rho,t}, \mathbf{x}_t) = \begin{cases} H(\mathbf{m}_{\rho,t}, \mathbf{x}_t), & d_{\mathcal{M}}(\mathbf{m}_{\rho,t}, \mathbf{x}_t) \leq A, \\ \frac{A}{d_{\mathcal{M}}(\mathbf{m}_{\rho,t}, \mathbf{x}_t)} H(\mathbf{m}_{\rho,t}, \mathbf{x}_t), & d_{\mathcal{M}}(\mathbf{m}_{\rho,t}, \mathbf{x}_t) > A. \end{cases} \quad (7)$$

This result indicates that the stochastic gradient matches that of the Karcher mean estimator when data points are close to the current centroid estimator (i.e., $d_{\mathcal{M}}(\mathbf{m}_{\rho,t}, \mathbf{x}_t) \leq A$), but it is scaled down for distant ones (i.e., $d_{\mathcal{M}}(\mathbf{m}_{\rho,t}, \mathbf{x}_t) > A$). Thus, the gradient becomes more robust against outliers and, consequently, less sensitive to abrupt changes in the online setting.

3.3. An adaptive CPD statistic

Using these two estimates \mathbf{m}_t and $\mathbf{m}_{\rho,t}$, we can define an adaptive CPD statistic by comparing them with geodesic distance on \mathcal{M} as:

$$g_t = d_{\mathcal{M}}(\mathbf{m}_t, \mathbf{m}_{\rho,t}). \quad (8)$$

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

The statistic g_t can be interpreted as a metric for detecting abrupt changes in streaming data on the manifold \mathcal{M} . Specifically, \mathbf{m}_t and $\mathbf{m}_{\rho,t}$ serve as adaptive estimators tracking the standard and robust centroids, respectively.

When a change point occurs within sequence $\{1, \dots, T\}$, the underlying data distribution shifts, causing measurements $\{x_1, \dots, x_T\}$ to include data points drawn from a different distribution $P_2(\mathbf{x})$ which can be viewed as outliers. The non-robust estimator \mathbf{m}_{t+1} adapts gradually to this new regime. In contrast, the robust estimator $\mathbf{m}_{\rho,t+1}$ remains stable against

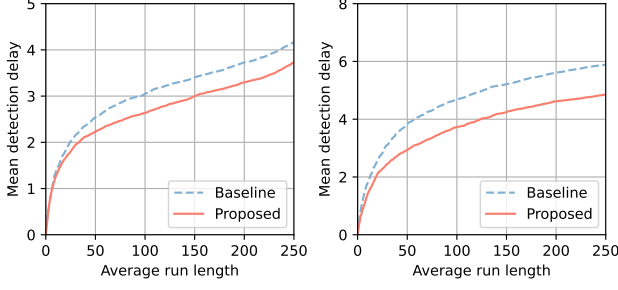


Fig. 1. ARL versus MDD for the compared algorithms on synthetic data on both \mathcal{S}_p^{++} (left) and \mathcal{G}_p^k (right)..

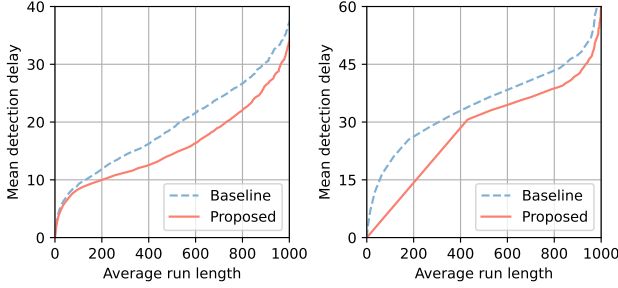


Fig. 2. ARL versus MDD for the compared algorithms on real data on both \mathcal{S}_p^{++} (left) and \mathcal{G}_p^k (right).

transient shifts and exhibits significant adaptation only when the contamination surpasses its robustness threshold. This behavior highlights the benefit of robust centroid estimation for Riemannian CPD, as lower sensitivity to abrupt changes makes the robust estimator an ideal baseline in constructing effective test statistic.

4. APPLICATIONS AND EXPERIMENTS

In this section, we apply Algorithm 1 to tackle two common instances of manifolds: the SPD manifold of $p \times p$ matrices, denoted by \mathcal{S}_p^{++} , and the Grassmann manifold, representing the set of k -dimensional linear subspaces of \mathbb{R}^p , denoted by \mathcal{G}_p^k . The implementation of Algorithm 1 on these manifolds requires manifold-specific definitions of the geodesic distance $d_{\mathcal{M}}(\mathbf{m}, \mathbf{x})$, the stochastic Riemannian gradients $H(\mathbf{m}, \mathbf{x})$ and $H_{\rho}(\mathbf{m}, \mathbf{x})$, and a second-order retraction operator $R_{\mathbf{m}}$. The definitions of $d_{\mathcal{M}}(\mathbf{m}, \mathbf{x})$, $H(\mathbf{m}, \mathbf{x})$ and $R_{\mathbf{m}}$ for \mathcal{S}_p^{++} and \mathcal{G}_p^k are provided in Section 5 of [15]. The corresponding robust gradient $H_{\rho}(\mathbf{m}, \mathbf{x})$ can be derived using the relation defined in (7) based on $H(\mathbf{m}, \mathbf{x})$.

We shall now illustrate the performance of Algorithm 1 for Riemannian CPD on both manifolds \mathcal{S}_p^{++} and \mathcal{G}_p^k . For Algorithm 1, the step size was set to $\alpha = 0.05$, and the parameter A was set to 1 for \mathcal{S}_p^{++} and 0.05 for \mathcal{G}_p^k . We selected a non-parametric online algorithm for Riemannian CPD [14, 15] as a baseline for comparison with Algorithm 1. To compare the

detection performance of these algorithms, 10^4 Monte Carlo simulations were performed to estimate the MDD and ARL for both synthetic and real data. Open-source code to reproduce the results is available on https://github.com/xiuheng-wang/CPD_RCE_release.

4.1. Experiment with synthetic data:

We first present results obtained from synthetically generated sequences of i.i.d. data on \mathcal{S}_p^{++} and \mathcal{G}_p^k .

SPD Manifold: We sampled matrices $\Sigma_t \in \mathcal{S}_p^{++}$ with $p = 10$ from a Wishart distribution with scaling matrix V and $p+2$ degrees of freedom. A total of 2000 samples were generated, with a change point at $t_r = 1500$, where V was reset.

Grassmann Manifold: The data $\pi(U_t) \in \mathcal{G}_p^k$ with $p = 20$ and $k = 5$ was generated in two steps. First, we sampled matrices Z_t from a random matrix Gaussian distribution. Then, the orthonormal matrices U_t were obtained as the left singular vectors corresponding to the k largest singular values of Z_t . A total of 2000 samples was generated, with a change point at $t_r = 1500$, where the mean of the matrix Gaussian distribution of Z_t was reset.

Results: The MDD as a function of ARL for all methods is depicted in Fig. 1 for both manifolds. It is evident that Algorithm 1 results in a lower MDD for a fixed ARL when compared to the baseline.

4.2. Experiment with real data:

We now present results obtained from real data on both \mathcal{S}_p^{++} and \mathcal{G}_p^k , focusing on the task of detecting speech segments within noisy backgrounds. We first combined 4 seconds of real speech from the TIMIT database [25] with 15 seconds of background noise recorded in real street environments from the QUT-NOISE database [26], ensuring a Signal-to-Noise Ratio (SNR) of -3 dB. Next, we applied the Short-Time Fourier Transform (STFT) [27] to the resulting one-dimensional audio signal, thereby obtaining frequency-domain representations as a $d = 128$ dimensional time series, $s_t \in \mathbb{R}^d$. Finally, we averaged neighboring frequency channels of s_t to produce a downsampled representation with reduced dimensionality of 16 channels.

SPD Manifold: We generated data points $\Sigma_t \in \mathcal{S}_p^{++}$ with $p = 16$ by computing covariance matrices from sliding windows, each containing 32 consecutive samples.

Grassmann Manifold: We applied truncated singular value decomposition (SVD), retaining only the first singular vector ($k = 1$), to the samples within each sliding window. This yielded orthonormal matrices U_t that define the subspaces $\pi(U_t) \in \mathcal{G}_p^k$.

Results: The performance curves for all methods are represented in Fig. 2. It is important to emphasize that this problem setting is particularly challenging due to the complexity of real acoustic signals and the non-i.i.d. nature of the extracted features. Nevertheless, Algorithm 1 consistently demonstrates superior performance in terms of MDD versus ARL compared to the baseline methods.

5. CONCLUSION

In this paper, we introduced a robust centroid estimation approach for Riemannian CPD. We proposed an adaptive test statistic comparing two specialized robust centroid estimators computed via Riemannian SGD: one emphasizing adaptability to new data, and the other enhancing robustness against abrupt changes. Importantly, our CPD statistic requires selecting only a single step size, a crucial advantage given that Riemannian SGD convergence is guaranteed only within a limited step-size range. Experimental results on both SPD and Grassmann manifolds confirm that our method consistently outperforms the previous approach based on two Karcher mean estimators.

6. REFERENCES

- [1] D. Gajic, Z. Djurovic, J. Gligorijevic, S. Di Gennaro, and I. Savic-Gajic, "Detection of epileptiform activity in eeg signals based on time-frequency and non-linear analysis," *Frontiers in computational neuroscience*, vol. 9, pp. 38, 2015.
- [2] D. Rybach, C. Gollan, R. Schluter, and H. Ney, "Audio segmentation for speech recognition using segment features," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 4197–4200.
- [3] 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.
- [4] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [5] F. Gustafsson, "The marginalized likelihood ratio test for detecting abrupt changes," *IEEE Transactions on automatic control*, vol. 41, no. 1, pp. 66–78, 1996.
- [6] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, pp. 107299, 2020.
- [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] Z. Harchaoui, E. Moulines, and F. Bach, "Kernel change-point analysis," *Advances in neural information processing systems*, vol. 21, 2008.
- [11] A. Ferrari, C. Richard, A. Bourrier, and I. Bouchikhi, "Online change-point detection with kernels," *Pattern Recognition*, p. 109022, 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. Dubey and H.-G. Müller, "Fréchet change-point detection," *The Annals of Statistics*, vol. 48, no. 6, pp. 3312–3335, 2020.
- [14] X. Wang, R. Borsoi, and C. Richard, "Online change point detection on Riemannian manifolds with Karcher mean estimates," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 2033–2037.
- [15] X. Wang, R. A. Borsoi, and C. Richard, "Non-parametric online change point detection on Riemannian manifolds," in *International Conference on Machine Learning*. PMLR, 2024, pp. 50143–50162.
- [16] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.
- [17] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [18] D. E. Tyler, "A distribution-free m-estimator of multivariate scatter," *The annals of Statistics*, pp. 234–251, 1987.
- [19] I. Ilea, L. Bombrun, R. Terebes, M. Borda, and C. Germain, "An m-estimator for robust centroid estimation on the manifold of covariance matrices," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1255–1259, 2016.
- [20] A. Breloy, S. Kumar, Y. Sun, and D. P. Palomar, "Majorization-minimization on the stiefel manifold with application to robust sparse pca," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1507–1520, 2021.
- [21] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [22] N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, 2023.
- [23] S. Bonnabel, "Stochastic gradient descent on Riemannian manifolds," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013.
- [24] H. Zhang and S. Sra, "First-order methods for geodesically convex optimization," in *Conference on Learning Theory*, 2016, pp. 1617–1638.
- [25] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," *Linguistic Data Consortium*, 1993, 1993.
- [26] D. Dean, S. Sridharan, R. Vogt, and M. Mason, "The qut-noise-timit corpus for evaluation of voice activity detection algorithms," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association*. International Speech Communication Association, 2010, pp. 3110–3113.
- [27] L. Cohen, *Time-frequency analysis*, vol. 778, Prentice hall New Jersey, 1995.