Correlation-aware Online Change Point Detection

Chengyuan Deng*
cd751@rutgers.edu
Rutgers University
New Brunswick, New Jersey, USA

Zhengzhang Chen†
zchen@nec-labs.com
NEC Labs America
Princeton, New Jersey, USA

Xujiang Zhao
xuzhao@nec-labs.com
NEC Labs America
Princeton, New Jersey, USA

Haoyu Wang
haoyu@nec-labs.com
NEC Labs America
Princeton, New Jersey, USA

Junxiang Wang
junwang@nec-labs.com
NEC Labs America
Princeton, New Jersey, USA

Jie Gao
jg1555@cs.rutgers.edu
Rutgers University
New Brunswick, New Jersey, USA

Haifeng Chen
haifeng@nec-labs.com
NEC Labs America
Princeton, New Jersey, USA

Abstract

Change point detection aims to identify abrupt shifts occurring at multiple points within a data sequence. This task becomes particularly challenging in the online setting, where different types of change can occur, including shifts in both the marginal and joint distributions of the data. In this paper, we address these challenges by tracking the Riemannian geometry of correlation matrices, allowing Riemannian metrics to compute the geodesic distance as an accurate measure of correlation dynamics.

We introduce Rio-CPD, a correlation-aware online change point detection framework that integrates the Riemannian geometry of the manifold of symmetric positive definite matrices with the cumulative sum (CUSUM) statistic for detecting change points. Rio-CPD employs a novel CUSUM design by computing the geodesic distance between current observations and the Fréchet mean of prior observations. With appropriate choices of Riemannian metrics, Rio-CPD offers a simple yet effective and computationally efficient algorithm. We also provide a theoretical analysis on standard metrics for change point detection within Rio-CPD. Experimental results on both synthetic and real-world datasets demonstrate that Rio-CPD outperforms existing methods on detection accuracy, average detection delay, and efficiency.

CCS Concepts

• **Computing methodologies** → **Machine learning algorithms**; **Temporal reasoning**; • **Information systems** → **Data mining**.

Keywords

Change point detection, Riemannian geometry, CUSUM, AIOps

*Work done during an internship at NEC Labs America.
†Corresponding author.

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3761021

ACM Reference Format:
Chengyuan Deng, Zhengzhang Chen, Xujiang Zhao, Haoyu Wang, Junxiang Wang, Jie Gao, and Haifeng Chen. 2025. Correlation-aware Online Change Point Detection. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3746252.3761021

1 Introduction

Change point detection (CPD) [42] aims to identify abrupt shifts in the underlying distribution of a system based on noisy temporal observations. CPD plays a vital role across various domains, including climatology [47], finance [44], and healthcare [62]. It can be performed in either an offline or online setting, with the latter posing greater challenges due to its real-time constraints and the need for minimal detection delay. In this work, we focus on unsupervised online CPD in a discrete-time setting, where multivariate time series are streamed, multiple change points may arise, and no ground-truth labels are available throughout the process.

Despite significant progress in online CPD techniques in recent years [4], efficiently capturing changes in various patterns, such as those in the marginal distribution (*e.g.*, independent magnitude) and joint distribution (*e.g.*, correlations between covariates), remains challenging. Correlation-aware CPD methods [8, 10, 30, 63, 64] have gained attention due to their practical impact in fields such as behavioral science [40] and root cause analysis in AIOps [54–56, 65, 66]. However, these methods were not originally designed for online use. The reason comes in two-fold: first, extracting correlations for CPD in multivariate time series often involves complex techniques like graph neural networks or probabilistic graphical models, which are inherently challenging to adapt to an online setting. Second, these methods typically require an extended processing time, making them inefficient for real-time applications. As a result, despite the pressing need, there remains limited work specifically tailored for correlation-aware online CPD.

On the other hand, a significant portion of the existing work on online CPD falls under the category of *Subspace Models*. These approaches estimate a subspace for each batch of time series data and then use a metric to compute the distance between consecutive

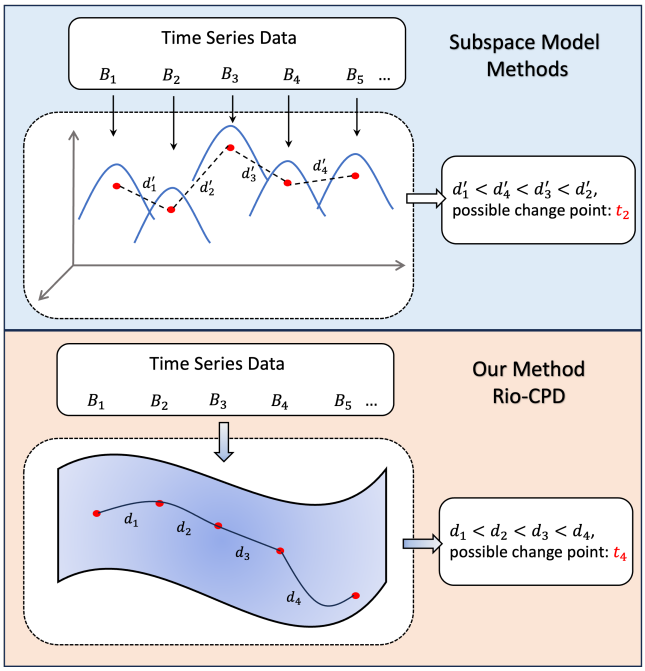

Figure 1: Comparison between Rio-CPD and subspace model. B_i represents the i-th batch of time series data, and d_i denotes the distances between neighboring batches in the subspace.

batches, as illustrated in Figure 1. A change point is detected if there is an abrupt change in this distance. Examples of subspace estimation techniques include singular spectrum analysis [4], state space models [21], dimension reduction [26], and refined metric design [28]. Although subspace models are relatively easy to adapt to online settings, they often suffer from accuracy issues due to distortions in subspace approximation, as well as computational inefficiencies stemming from their high processing cost.

To address these challenges, in this work, we extend the subspace model and propose Rio-CPD, a Riemannian geometry-based non-parametric method for correlation-aware Online Change Point Detection. At a schematic level, Rio-CPD is inspired by the observation that Pearson correlation matrices are symmetric positive semi-definite, hence can be characterized by Riemannian geometry and all fall into one subspace of Riemannian manifold. The distance between the correlations of two batches of time series data corresponds to the geodesic between two points on the Riemannian manifold, which can be computed using a specific Riemannian metric. These distances are then used to construct the cumulative sum (CUSUM) statistic, which performs sequential hypothesis test to determine whether a time step is a change point. The Rio-CPD framework is flexible with the choice of Riemannian metrics and CUSUM statistic construction, although there may be trade-offs. We employ the Log-Euclidean and Log-Cholesky metrics, which avoid Riemannian optimization in CUSUM statistic computation, thereby ensuring efficiency. Details of our design are provided in Section 3.

We present a comparison between Rio-CPD and subspace model-based methods in Figure 1. Rio-CPD addresses key limitations of subspace models by: (1) operating directly on the Riemannian manifold of correlation matrices rather than an estimated subspace, which can introduce distortion; and (2) bypassing the time-consuming subspace learning process. As a result, the Rio-CPD framework enhances the subspace model with improved accuracy and computational efficiency.

Our contributions can be summarized as follows:

- ***Problem***: We tackle the problem of correlation-aware online change point detection, aiming to accurately and efficiently identify changes in both individual variables and correlations within multivariate time series data. Existing methods either lack the ability to handle correlations or are unsuitable for real-time detection, motivating the need for a new approach.
- ***Framework***: We propose Rio-CPD, a non-parametric framework that leverages the Riemannian geometry of correlation matrices and the CUSUM procedure to detect change points. Rio-CPD is capable of detecting changes in both independent and correlated patterns while being highly efficient and flexible in choosing Riemannian metrics to balance performance trade-offs. We also provide a *theoretical analysis* on two standard metrics for change point detection of Rio-CPD.
- ***Evaluations***: We conduct extensive experiments on both synthetic and real-world datasets to validate the effectiveness of our approach. The results show that Rio-CPD significantly outperforms the state-of-the-art methods in terms of detection accuracy and efficiency.

2 Preliminary

2.1 Problem Statement

Let $X(t) = [X_1(t), \ldots, X_m(t)] \in \mathbb{R}^m$ be an observation of a discrete multivariate time series at time index t, where $t \in [T]$, and denote $f_i : \mathbb{Z}^+ \to \mathbb{R}$ as the latent distribution of the i-th time series, such that the observation at time t take the form of $X(t) = f(t) + e(t)$, where $f(t) = [f_1(t), \ldots, f_m(t)]$ and $e(t)$ is a zero-mean i.i.d. random variable representing the noise. In particular, the change point detection (CPD) task aims to find all $\tau \in [T]$ such that:

$$X(t) = \begin{cases} f(t) + e(t), & t < \tau \\ f'(t) + e(t), & t \geq \tau \end{cases}$$

for some functions $f \neq f'$. Another perspective of the CPD problem is to perform sequential hypothesis testing at each time step t such that one of the following is accepted:

$$H_0 : \mathbb{E}[X(t)] = f(t)$$
$$H_1 : \mathbb{E}[X(t)] = f'(t)$$

Let $\hat{\tau} = \inf\{t | H_1 \text{ is accepted at } t\}$, for any CPD algorithm, an essential property is to identify the change point promptly, *i.e.*, $\hat{\tau} > \tau$ and $\hat{\tau} - \tau$ is small.

2.2 Standard Notions in Riemannian Geometry

A manifold $\mathcal{M}$ is a topological space that is locally diffeomorphic to a Euclidean space. To measure geometric properties, a Riemannian manifold is defined as a manifold equipped with a Riemannian metric, formally as below:

Definition 2.1 (Riemannian Manifold)**.** A Riemannian manifold $(\mathcal{M}, g)$ is a smoothed manifold $\mathcal{M}$ endowed with a smoothly varying family of inner products $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$, where $T_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$.

The geodesic of a Riemannian manifold can be defined by using the inner product. Specifically, for two points $p, q \in \mathcal{M}$, the geodesic distance between them follows:

$$d_{\mathcal{M}}(p, q) = \inf_{\gamma} \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))}\, dt, \tag{1}$$

where $\gamma : [0, 1] \to \mathcal{M}$ is a smooth curve such that $\gamma(0) = p$ and $\gamma(1) = q$.

To facilitate computations on such manifolds, the exponential and logarithm maps play a crucial role by linking the geometry of a Lie group with its Lie algebra. The Lie algebra can be seen as the tangent space of the Lie group at the identity element, providing a local linear approximation of the manifold. In our context, these maps establish a one-to-one correspondence between symmetric positive definite (SPD) matrices and a vector space structure, ensuring that geodesics are well-defined and can be computed effectively.

Definition 2.2 (Exponential and Logarithm Map of Matrices)**.** Given a matrix A, the exponential map is defined by $\exp(A) := \sum_{k=0}^{\infty} A^k/k!$, while the logarithm map, denoted as $\log(A)$, serves as its inverse.

Finally, we introduce the Fréchet mean of a set of matrices within a given metric space.

Definition 2.3 (Fréchet mean)**.** Let $(\mathcal{M}, d)$ be a complete metric space, and let $P_1, \dots P_n$ be points in $\mathcal{M}$. The Fréchet mean of $P_1, \dots P_n$ is defined as:

$$\sigma_{\mathcal{M}} = \operatorname*{argmin}_{x \in \mathcal{M}} \sum_{i=1}^{n} d^2(x, P_i) \tag{2}$$

3 Methodology

The proposed framework assumes linear dependencies between different attributes of the time series. Our approach is therefore motivated by the observation that in practice[1], Pearson correlation matrices lie in $\mathcal{S}_+^n$, which exhibits Riemannian structures [6]. This allows us to apply a Riemannian metric to correlation matrices with minimal distortion. Based on this metric, the Rio-CPD framework proceeds in three stages.

First, the Riemannian metric is used to track the distances between consecutive correlation matrices until a change point is identified. Next, we construct the CUSUM statistics by measuring the distance between the current correlation matrix and the "centroid" of the sub-manifold representing the collection of past correlation matrices. If a change point is present, this distance will be significantly large, indicating that the current matrix should belong to the cluster representing the post-change distribution, and is an "outlier" to the current one. Finally, the third stage involves applying the CUSUM sequential hypothesis test to detect and report a change point as soon as it occurs. An illustration of the Rio-CPD framework is provided in Figure 2.

[1]Correlation matrices are positive semi-definite, but under reasonable assumptions, such as linear independence of data columns, they are positive definite.

It is important to note that constructing the CUSUM statistics is non-trivial, as it is closely related to the choice of Riemannian metrics. In general, the centroid of a Riemannian manifold, represented by the Fréchet mean, may not have a closed-form solution and often requires approximation through optimization. However, with the use of two specific Riemannian metrics, known as the Log-Euclidean and Log-Cholesky metrics, the Fréchet mean has a closed-form solution. This property allows Rio-CPD to bypass the complexities of Riemannian optimization.

3.1 CUSUM Procedure

The CUSUM statistic is a measure of the likelihood that a given timestamp represents a change point, and it has been widely used in hypothesis testing. To construct the CUSUM statistic at timestamp t, a detection score is required, which typically evaluates the "distance" between the observation at timestamp t and the distribution of the time series observed so far. Formally,

Definition 3.1 (CUSUM Statistic)**.** Given an observation $X(t)$ at time step t and a detection score $D(t)$, the CUSUM statistic for the time series $\{X(1), \dots, X(t)\}$ is defined as $y(t) = \max_{1 \le i \le t} \sum_{j=i}^{t} D(j)$.

It is well-known from the literature [42] that Theorem 3.1 can alternatively be expressed as $y(t) = \max\{y(t-1) + D(t), 0\}$, where $y(0) = 0$. This recursive formulation improves the efficiency of testing by utilizing only the previous step's history. The CUSUM procedure proceeds as follows: At each timestamp t, we perform the hypothesis test:

$$\begin{aligned} H(t) = H_{\mathbb{1}_{y(t) \ge \rho}}, \text{ where} \\ H_0 : t \text{ is not a change point,} \\ H_1 : t \text{ is a change point.} \end{aligned}$$

Note that $\rho > 0$ is a threshold. Therefore, the set of change points is given by $t \mid y(t) \ge \rho$.

Although the CUSUM procedure may seem straightforward for change point detection, designing the detection score $D(t)$ is often challenging and plays a critical role in the overall performance of the method. In parametric settings, $D(t)$ can be defined as the log-likelihood ratio between two distributions. However, this requires knowledge of the post-change distribution, which is often unavailable in practice and limits its applicability to general scenarios.

In the following sections, we explain how Rio-CPD constructs $D(t)$. A key requirement for $D(t)$ is its ability to effectively differentiate between change points and regular timestamps, which necessitates that $\mathbb{E}[D(t)|H_0] < 0$ and $\mathbb{E}[D(t)|H_1] > 0$.

3.2 Log-Euclidean Metric

We first introduce the Lie group structure of $\mathcal{S}_+^n$, which enables a bi-invariant Riemannian metric derived from any inner product between $\mathcal{S}_+^n$ and its tangent space. Given two matrices $P_1, P_2 \in \mathcal{S}_+^n$, we define the logarithm addition and multiplication, with the exponential and logarithm map (see Definition 2.2).

$$P_1 \oplus P_2 = \exp(\log P_1 + \log P_2)$$
$$\lambda \odot P_1 = \exp(\lambda \cdot \log P_1)$$

where λ is a real number.

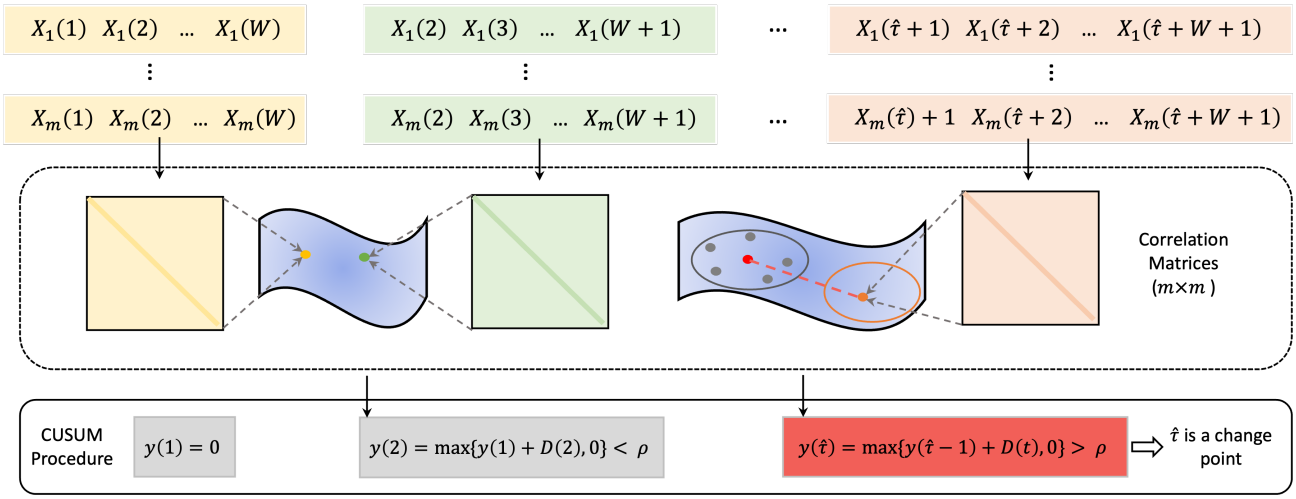

Figure 2: An overview of the proposed CPD framework: (1) construct a correlation matrix for each batch within the sliding window, (2) compute the distance of the current observation from the previous centroid (large distances suggest a new cluster), and (3) compute the CUSUM statistic using these distances to perform a hypothesis test for change detection.

The matrix logarithm for Log-Euclidean is defined as:

$$\phi_{LE}(P_1) = U \ln(\Sigma) U^T$$

where $P_1 = U\Sigma U^T$ is the eigenvalue decomposition.

The geodesic distance under the Log-Euclidean metric is:

$$d_{LE}(P_1, P_2) = \|\phi_{LE}(P_1) - \phi_{LE}(P_2)\|_F \tag{3}$$

While the computation of the Log-Euclidean metric can be theoretically demanding due to the logarithm map, there are techniques available to reduce computation time. One notable advantage of the Log-Euclidean metric is that it provides a closed-form solution for the Fréchet mean of SPD matrices, which can be regarded as the "centroid" of the $\mathcal{S}_+^n$ geometry. Similar to the centroid in Euclidean space (*e.g.*, the k-means objective), the Fréchet mean minimizes the squared error of the geodesic distances to all SPD matrices.

The Fréchet mean of Log-Euclidean geometry (also known as Log-Euclidean mean) is given by [6]:

$$\sigma_{LE}(P_1, \ldots, P_n) = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \phi_{LE}(P_i)\right) \tag{4}$$

It is worth noting that the Log-Euclidean mean is a natural generalization of the geometric mean. If P_i are positive real numbers, their geometric mean follows the same formula. The Log-Euclidean and Log-Cholesky metrics are the only Riemannian metrics that admit a closed-form solution for the Fréchet mean. For other metrics, the Fréchet mean may not even be unique.

3.3 Log-Cholesky Metric

The Log-Euclidean and Log-Cholesky metrics [35] share several similarities and, at a high level, are topologically equivalent because $\mathcal{L}^n \simeq \mathcal{S}^n \simeq \mathbb{R}^{n(n+1)/2}$ [2]. The Log-Euclidean metric is obtained through the matrix logarithm: $\mathcal{S}_+^n \to \mathcal{S}^n$, while the Log-Cholesky metric is based on $\mathcal{S}_+^n \to \mathcal{L}^n$. The definition of the Log-Cholesky metric begins with the Cholesky decomposition, which provides a unique representation of any Hermitian and positive definite matrix A in the form LL^T, where L is a lower triangular matrix and L^T is its upper triangular counterpart. The matrix logarithm for the Log-Cholesky geometry is defined as:

$$\phi_{LC}(P_1) = \lfloor L \rfloor + \ln(\mathbb{D}(L))$$

where $\lfloor L \rfloor$ is the strictly lower triangular part of L, and $\mathbb{D}(L)$ denotes the diagonal matrix extracted from L, with all other entries set to 0.

The geodesic distance under Log-Cholesky metric is defined as:

$$\begin{aligned} d_{LC}(P_1, P_2) &= \|\phi_{LC}(P_1) - \phi_{LC}(P_2)\|_F \\ &= \Big(\|\lfloor L_1 \rfloor - \lfloor L_2 \rfloor\|_F^2 \\ &\quad + \|\ln(\mathbb{D}(L_1)) - \ln(\mathbb{D}(L_2))\|_F^2 \Big)^{1/2} \end{aligned} \tag{5}$$

where $P_1 = L_1 L_1^T$ and $P_2 = L_2 L_2^T$.

The Log-Cholesky metric also has a convenient closed-form expression for its Fréchet mean over n matrices $P_1, \ldots, P_n \in \mathcal{S}_+^n$:

$$\sigma_{LC}(P_1, \ldots, P_n) = \sigma_L(L_1, \ldots, L_n) \cdot \sigma_L(L_1, \ldots, L_n)^T$$

where P_i has Cholesky decomposition $L_i L_i^T$ and σ_L is the Fréchet mean of Cholesky space $\mathcal{L}_+^n$:

$$\sigma_L(P_1, \ldots, P_n) = \frac{1}{n}\sum_{i=1}^{n} \lfloor L_i \rfloor + \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log \mathbb{D}(L_i)\right) \tag{6}$$

3.4 RIO-CPD Algorithm

RIO-CPD requires one parameter to be specified in advance: the size of the sliding window W. The algorithm proceeds in three steps at each time step t: (1) Using the time series data from t to $t + W - 1$, construct the correlation matrix $\mathbf{B}_t$. (2) Monitor the

[2] Notation: $\mathcal{L}^n$ denotes the Cholesky space of dimension n, and $\simeq$ represents homotopic equivalence. These notations are used only once here.

geodesic distance between $\mathbf{B}_t$ and σ_{t-1}, the Fréchet mean of the collection of previous correlation matrices. (3) Compute the CUSUM statistic based on this distance and perform hypothesis testing. If the test does not detect a change point, proceed to $t+1$; otherwise, report a change point at t and repeat the process for the next batch.

Step 1. Transform observations in the sliding window into correlation matrices. Suppose at time t we have $\mathbf{X}_t = [X(t), \ldots, X(t+W-1)]$, then construct

$$\mathbf{B}_t = \tilde{\mathbf{X}}_t \cdot \tilde{\mathbf{X}}_t^T, \text{where } \tilde{\mathbf{X}}_t = \frac{\mathbf{X}_t - \mathbb{E}[\mathbf{X}_t]}{Var(\mathbf{X}_t)} \text{ is normalized.}$$

Step 2. Let $\mathfrak{g}$ be either the Log-Euclidean or Log-Cholesky metric. We compute $\sigma^{\mathfrak{g}}_{t-1}$, the Fréchet mean of $\mathbf{B}_1, \ldots, \mathbf{B}_{t-1}$ using Equation (4) or Equation (6). Next, we calculate the distance from $\mathbf{B}_t$ to $\sigma^{\mathfrak{g}}_{t-1}$, *i.e.*, $d_t = \mathfrak{g}(\mathbf{B}_t, \sigma^{\mathfrak{g}}_{t-1})$. [3]

Step 3. Construct the detection score $D(t)$ required for CUSUM. We define $D(t)$ as d_t minus r_{t-1}, the radius of the subspace formed by $\mathbf{B}_1, \ldots, \mathbf{B}_{t-1}$, i.e.

$$D(t) = d_t - r_{t-1} = \mathfrak{g}(\mathbf{B}_t, \sigma^{\mathfrak{g}}_{t-1}) - \max_{i \in [t-1]} \mathfrak{g}(\mathbf{B}_i, \sigma^{\mathfrak{g}}_{t-1})$$

With the detection score $D(t)$, the CUSUM test iteratively computes $y(t)$ for $t > W$, starting with $y(W) = 0$, until a change point is detected, *i.e.*, $y(t) > \rho$ for a given threshold ρ. Specifically,

$$\hat{\tau} = \inf_{t>W} \{t \mid y(t) > \rho\}.$$

The procedure can be extended to handle multiple change points by restarting the process at $\hat{\tau}+1$ after detecting a change point $\hat{\tau}$, with a new base correlation matrix $\mathbf{B}_{\hat{\tau}+1}$. We provide further insights into our algorithm and discuss its computational complexity in the remarks below.

Remark 3.2. The design of the detection score $D(t)$ is inspired by clustering and incorporates strong geometric insights. Essentially, Rio-CPD first projects the correlation dynamics into the Riemannian manifold, if the input data follows a stable distribution, the projection would become a cluster of points in the Riemannian manifold. Therefore, a change in distribution would lead to a shift of the cluster. Rio-CPD identifies t as a change point if $\mathbf{B}_t$ is an outlier from the cluster formed by $\mathbf{B}_1, \ldots, \mathbf{B}_{t-1}$. To detect an outlier, we calculate the difference between the current distance and the "radius" of the cluster. If this difference exceeds a certain threshold, it is likely that $\mathbf{B}_t$ is an outlier, indicating a change point.

Remark 3.3. The primary computational bottleneck for our algorithm arises from eigen decomposition (Log-Euclidean) and Cholesky decomposition (Log-Cholesky), both of which are performed with $O(m^3)$ time complexity for m-dimensional symmetric matrices. In theory, the exponent can be reduced to a constant close to ω (the fast matrix multiplication constant). Note that m is the number of time series in our setting and is typically a constant, making the algorithm highly efficient in practice.

[3]Here we consider a 1-lag. If the observations are sampled at high frequency, the algorithm can use $\mathbf{B}_1, \mathbf{B}_{1+L}, \mathbf{B}_{1+2L}, \ldots$, where $L \geq 1$ becomes a parameter.

4 Theoretical Analysis

There are two standard metrics for evaluating the performance of CPD algorithms: the Average Run Length (ARL) and the Expected Detection Delay (EDD). The ARL corresponds to the expected stopping time in the absence of any change points, while the EDD measures the delay in detecting a true change once it occurs. We present theoretical analyses of the proposed Rio-CPD framework under these two metrics, drawing on classical results from prior work [49] and the derivation of the Kullback–Leibler (KL) divergence with respect to Riemannian geometry.

In deriving our theoretical results, we introduce one key parameter. Although Rio-CPD is flexible with diverse choices of Riemannian metrics, we denote by δ that any Riemannian metric has distortion no larger than δ for Rio-CPD. Our main result is stated below.

Theorem 4.1. *Let $\mathbb{E}_\infty[\mathcal{T}_W]$ be ARL and $\mathbb{E}_0[\mathcal{T}_W]$ stand for EDD, suppose ρ is the minimum threshold appeared in the CUSUM sequential tests of* Rio-CPD*, we have:*

$$\mathbb{E}_\infty[\mathcal{T}_W] \leq \frac{8\delta^2 \cdot e^\rho}{\rho^2}, \mathbb{E}_0[\mathcal{T}_W] \leq \frac{8\delta^2}{\rho}$$

Theorem 4.1 provides upper bounds on both ARL and EDD, as worst-case performance guarantees for Rio-CPD. Since our algorithm consistently employs the three-sigma rule for selecting ρ, its performance is primarily constrained by the potential distortion of the chosen Riemannian metrics, which may arise from metric design, selection, or noise in the time-series data.

Proof. We start with the textbook results restated from [49].

Proposition 4.2. *Let ρ be the threshold of CUSUM test, f_0 and f_∞ be the pre- and post-change distribution of the time series, then we have the following:*

$$\mathbb{E}_\infty[\mathcal{T}_W] = \frac{e^\rho}{\mathbb{E}_\infty[\log(f_\infty(x)/f_0(x))]}(1+o(1)) \tag{7}$$

$$\mathbb{E}_0[\mathcal{T}_W] = \frac{\rho}{\mathbb{E}_0[\log(f_0(x)/f_\infty(x))]}(1+o(1)) \tag{8}$$

Recall the definition of Kullback-Leibler divergence between two distributions $f(x), g(x)$:

$$D_{\mathrm{KL}}(f||g) = \int_x f(x) \log \frac{f(x)}{g(x)} dx \tag{9}$$

Therefore, above equations for ARL and EDD can be rewritten w.r.t. the KL divergence:

$$\mathbb{E}_\infty[\mathcal{T}_W] = \frac{e^\rho}{D_{\mathrm{KL}}(f_\infty||f_0)}(1+o(1)) \tag{10}$$

$$\mathbb{E}_0[\mathcal{T}_W] = \frac{\rho}{D_{\mathrm{KL}}(f_0||f_\infty)}(1+o(1)) \tag{11}$$

We next establish a connection between the KL divergence and Riemannian metrics. In particular, we show that

$$\mathfrak{g}^{FI}(\theta_\infty, \theta_0) = \sqrt{2D_{\mathrm{KL}}(f_\infty||f_0)} \tag{12}$$

where $\mathfrak{g}^{FI}$ refers to the Fisher information metric, which is a Riemannian metric. θ_∞, θ_0 are two points on the Riemannian manifold corresponding to f_∞, f_0. This result has been discussed in prior works without explicit proof, for completeness we show the details.

Now we prove that Equation (12) is true, starting from defining the Fisher Information matrix.

Definition 4.3. Given a manifold with coordinates $\Theta = (\theta_1, \theta_2, \dots)$, let $p(x|\theta)$ be the likelihood and normalized to 1, i.e. $\int_x p(x|\theta)dx = 1$.

$$\mathfrak{g}_{ij}^{FI}(\theta) = \int_x \frac{\partial^2 \log p(x|\theta)}{\partial\theta_i \partial\theta_j} p(x|\theta)dx$$

Our goal is to show $D_{\mathrm{KL}}(\theta_i||\theta_j) = \frac{1}{2}(\theta_i - \theta_j)^T \cdot \mathfrak{g}_{ij}^{FI}(\theta) \cdot (\theta_i - \theta_j)$. It is essentially obtained by Taylor expansion.

$$\begin{aligned} D_{\mathrm{KL}}(\theta_i \parallel \theta_j) \approx\ & D_{\mathrm{KL}}(\theta_i \parallel \theta_j)\big|_{\theta_j=\theta_i} \\ & + (\theta_j - \theta_i)^T \frac{\partial D_{\mathrm{KL}}(\theta_i \parallel \theta_j)}{\partial\theta_j}\Big|_{\theta_j=\theta_i} \\ & + \frac{1}{2}(\theta_j - \theta_i)^T \frac{\partial^2 D_{\mathrm{KL}}(\theta_i \parallel \theta_j)}{\partial\theta_j \partial\theta_j}\Big|_{\theta_j=\theta_i} (\theta_j - \theta_i) \end{aligned}$$

It is convenient to verify that the first two terms are 0.

$$\begin{aligned} D_{\mathrm{KL}}(\theta_i \parallel \theta_j)\big|_{\theta_j=\theta_i} &= \int p(x;\theta_i) \log \frac{p(x;\theta_i)}{p(x;\theta_j)} dx\Big|_{\theta_j=\theta_i} \\ &= \int p(x;\theta_i) \log 1 dx = 0 \end{aligned}$$

$$\begin{aligned} &\frac{\partial D_{\mathrm{KL}}(\theta_i \parallel \theta_j)}{\partial\theta_j}\Big|_{\theta_j=\theta_i} = \frac{\partial}{\partial\theta_j} \int p(x;\theta_i) \log \frac{p(x;\theta_i)}{p(x;\theta_j)} dx\Big|_{\theta_j=\theta_i} \\ &= \frac{\partial}{\partial\theta_j} \int p(x;\theta_i) \left(\log p(x;\theta_i) - \log p(x;\theta_j)\right) dx\Big|_{\theta_j=\theta_i} \\ &= -\int p(x;\theta_i) \frac{\partial}{\partial\theta_j} p(x;\theta_j) \frac{1}{p(x;\theta_j)} dx\Big|_{\theta_j=\theta_i} \\ &= -\frac{\partial}{\partial\theta_i} \int p(x;\theta_i) dx = 0 \end{aligned}$$

It then falls into the fact that the third term is related to the Fisher Information metric.

$$\begin{aligned} &\frac{\partial^2 D_{\mathrm{KL}}(\theta_i \parallel \theta_j)}{\partial\theta_j \partial\theta_j}\Big|_{\theta_j=\theta_i} \\ &= \frac{\partial^2}{\partial\theta_j \partial\theta_j} \int p(x;\theta_i) \left(\log p(x;\theta_i) - \log p(x;\theta_j)\right) dx\Big|_{\theta_j=\theta_i} \\ &= -\int p(x;\theta_i) \frac{\partial^2}{\partial\theta_j \partial\theta_j} \log p(x;\theta_j) dx\Big|_{\theta_j=\theta_i} \\ &= -\int p(x;\theta_i) \frac{\partial^2}{\partial\theta_i \partial\theta_i} \log p(x;\theta_i) dx = \mathfrak{g}_{ij}^{FI}(\theta) \end{aligned}$$

We have verified Equation (12). Recall our CUSUM statistics design in Rio-CPD, the hypothesis testing reports a change point when $D(t) = d\mathfrak{g}(B_t, c_{t-1}) \geq \rho$, where $D(t)$ is the distance between the current batch and the centroid of previous points, and $\mathfrak{g}$ is a Riemannian metric used for Rio-CPD. With the assumption that Riemannian metrics have distortion at most δ, we have

$$\begin{aligned} D(t) = \mathfrak{g}(B_t, c_{t-1}) &\leq \mathfrak{g}^{FI}(\theta_\infty, \theta_0) \cdot 2\delta \\ &\approx 2\sqrt{2D_{\mathrm{KL}}(f_\infty||f_0)}\delta \end{aligned}$$

Plugging this back into Theorem 4.2, we obtain upper bounds for ARL and EDD:

$$\mathbb{E}_\infty[\mathcal{T}_W] \leq \frac{8\delta^2 \cdot e^\rho}{\rho^2} \tag{13}$$

$$\mathbb{E}_0[\mathcal{T}_W] \leq \frac{8\delta^2}{\rho} \tag{14}$$

Our proof does not rely on the assumptions with respect to the geometric parameters of a Riemannian manifold such as geodesic convexity, Lipschitz gradient, smoothness, etc. However, the assumption that Riemannian metrics have small distortion is also non-trivial. we refer to future work to directly establish a connection between the KL divergence and other Riemannian metrics, especially the Log-Euclidean and Log-Cholesky metrics used in this paper. Another future direction of theoretical interest is to characterize the distribution changes in correlation matrices and show how it might impact the performance of Rio-CPD. □

5 Experiments

In this section, we evaluate the performance of Rio-CPD in terms of accuracy and efficiency. We begin by evaluating general scenarios in which the multivariate time series are derived from a dynamic system, with various types of change points. We then consider the task of human action recognition as a concrete application of change point detection and present the corresponding results. All experiments were conducted on a Linux system equipped with AMD EPYC 7302 16-core processors and two Nvidia RTX5500 GPUs with 24GB of memory, although most baseline methods do not require the use of a GPU.

5.1 Setup

Datasets. We evaluate our model Rio-CPD using both synthetic and real-world datasets. Three synthetic datasets are derived from a particle-spring system [64], which features five particles moving within a rectangular space. The particles are randomly connected by invisible springs, and their movement follows Newton's and Hooke's laws. The *Connection* data set exhibits correlation changes, with change points occurring when spring connections change. In contrast, the *Speed* and *Location* datasets show change points related to the particles' speed and location, respectively. For *real-world data*, we use several benchmark datasets commonly compared in prior works, including *Beedance*[4] and *HASC*[5], which do not necessarily exhibit correlation-based change points. In addition, we include the *Product Review Microservice* dataset [67], which contains change points due to system irregularities. In this dataset, the system performance metrics are correlated because of the underlying system architecture. We provide basic statistics for all datasets inTable 1.

Baselines. We compare Rio-CPD with four baselines representing different families of online change point detection methods: **KL-CPD** [12], which extends the kernel two-sample test and optimizes a lower bound of the test power via an auxiliary generative model; **BOCPDMS**, the Spatial-temporal Bayesian Online CPD [29], which augments the vanilla Bayesian CPD with Bayesian vector autoregressions; **MSSA-CPD** [4], which employs CUSUM statistics

[4]https://sites.cc.gatech.edu/ borg/ijcv_pssIds/
[5]http://hasc.jp/hc2011/

Table 1: Dataset statistics. #CP represents the number of change points, and #TS denotes the number of independent time series (not concatenated). A checkmark in the Correlation column indicates that the change point is due to correlation changes.

Dataset	Length	#CP	#TS	Dimension	Correlation
Connection	100	1	50	$\mathbb{R}^5$	✓
Speed	100	1	50	$\mathbb{R}^5$	✗
Location	100	1	50	$\mathbb{R}^5$	✗
Microservice	1548-1767	8	4	$\mathbb{R}^6$	✓
Beedance	608-1124	117	6	$\mathbb{R}^3$	✗
HASC	11738-12000	196	18	$\mathbb{R}^3$	✗

based on subspace estimation via multivariate Singular Spectrum Analysis; and **Contra-CPD** [45], which designs a test statistic that extends the concept of maximizing the discrepancy between pre- and post-change distributions. For Contra-CPD, we select the polynomial function family among its variants.

Evaluation metrics. We evaluate detection performance primarily using the *F1 score*, which is crucial for accuracy, as CPD can be treated as a binary classification problem. The second metric is *detection delay*, which is particularly important in practice when delays can have significant consequences. The delay is reported by counting the number of time steps. We also report *running time* to assess efficiency.

Parameters. Only two parameters are required: the threshold ρ of the CUSUM statistic and the window size W. We adopt a consistent strategy for selecting ρ, the three-sigma rule [46], namely $\rho = 3 \cdot \sigma$, where σ is the standard deviation calculated over the data samples until a change point occurs. The three-sigma rule is a well-established statistical concept and a common heuristic widely used across various applications. In our experiments, W is usually set as 10 or 20 and we did not prioritize extensive optimization of this parameter.

5.2 Detection Performance

We first evaluate the accuracy of Rio-CPD on both synthetic and real-world datasets and report the F1 score under two settings: *Default*, which uses the vanilla parameter initialization, and *Best*, which involves a fine-tuning process. A change point is considered successfully detected if it falls within the sliding window reported by the algorithm. For baseline methods that do not use a sliding window, we omit the error caused by the value of W to ensure a fair comparison.

The results in Table 2 indicate that Rio-CPD, with both Log-Euclidean (LE) and Log-Cholesky (LC) metrics, consistently performs better than or at least competitively with other methods. In particular, for datasets with correlation-based change points, Rio-CPD significantly outperforms other methods. We observe that other methods suffer from both false negatives and false positives—they tend to miss correlation changes while also reporting more change points due to local perturbations. Notably, KL-CPD has an F1 score of 0 on the Microservice dataset due to the absence of any true positive detections. For other datasets with change points not specified as correlation-based, Rio-CPD also shows favorable performance. This is due to the dual ability of Rio-CPD to detect both general and correlation-based change points. We provide more discussions on this observation with a toy example in Section 7.

Table 2: F1 score on real and synthetic datasets. A higher score indicates better performance.

Algorithm	Microservice		Beedance		HASC	
	Default	Best	Default	Best	Default	Best
KL-CPD	0	0	0.092	0.167	0.078	0.204
BOCPDMS	0.061	0.109	0.092	0.167	0.078	0.204
MSSA-CPD	0.154	0.308	0.500	**0.659**	0.177	0.327
Contra-CPD	0.122	0.258	0.167	0.293	0.151	0.239
Rio-CPD (**LE**)	0.778	0.875	0.518	0.625	0.320	0.345
Rio-CPD (**LC**)	**0.933**	**0.933**	**0.535**	0.643	**0.360**	**0.463**
	Connection		Speed		Location	
	Default	Best	Default	Best	Default	Best
KL-CPD	0.087	0.120	0.155	0.263	0.052	0.199
BOCPDMS	0.031	0.096	0	0.114	0	0.043
MSSA-CPD	0.179	0.330	0.292	0.485	0.153	0.308
Contra-CPD	0.268	0.304	0.315	0.396	0.283	0.342
Rio-CPD (**LE**)	0.446	0.496	0.412	**0.510**	0.378	**0.493**
Rio-CPD (**LC**)	**0.494**	**0.511**	**0.473**	0.500	**0.459**	0.482

Furthermore, the LC metric is claimed to be more numerically stable and computationally efficient compared to the LE metric [35], which is supported by our experiments, as Rio-CPD with the LC metric demonstrates slightly better performance. An example of change points detected by Rio-CPD is shown in Figure 3.

Table 3: Average delay on all datasets. 'N.A.' is used when no positive detection occurs within a reasonable delay window. All values are rounded to the nearest integer.

Algorithm	Microservice	Beedance	HASC
KL-CPD	11	7	N.A.
BOCPDMS	8	6	N.A.
MSSA-CPD	14	11	41
Contra-CPD	10	16	33
Rio-CPD (**LE**)	0	8	25
Rio-CPD (**LC**)	0	6	19
	Connection	Speed	Location
KL-CPD	N.A.	N.A.	N.A.
BOCPDMS	9	6	7
MSSA-CPD	2	4	3
Contra-CPD	5	5	4
Rio-CPD (**LE**)	2	2	2
Rio-CPD (**LC**)	2	2	3

We next evaluate the average detection delay for each algorithm. Detection delay measures the sensitivity of a CPD algorithm in an online setting, with a shorter delay preferable. To avoid false alarms, we calculate the average delay only for points detected within twice the window size. Detections outside this range are not considered related to the current change point. The results are presented in Table 3. Additionally, we demonstrate the running time of all algorithms on three datasets: Microservice, Beedance,

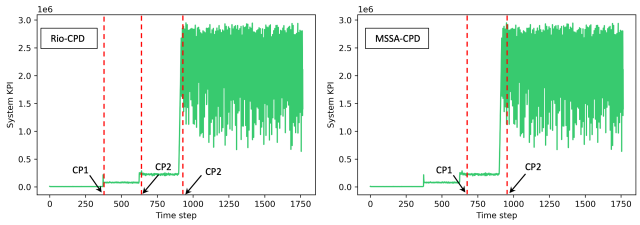

Figure 3: Detected change points by Rio-CPD and MSSA-CPD on the Microservice dataset, with false alarms omitted for MSSA-CPD.

and Connection in Figure 4. The reported numerical values are logarithmically transformed.

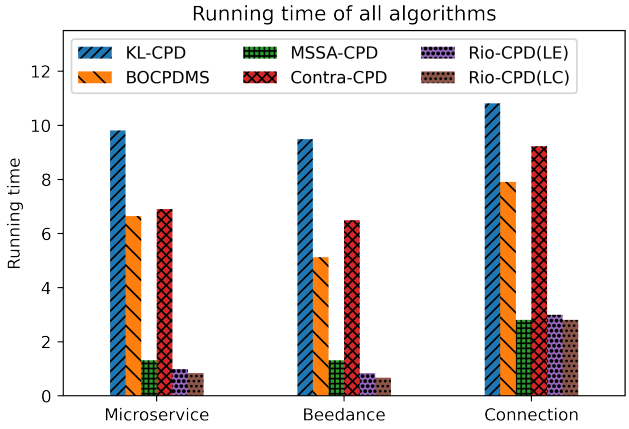

Figure 4: Running time comparison on three datasets.

An immediate observation is that Rio-CPD, with either metric, runs faster than all baseline methods and has a lower detection delay. This efficiency is due to the simplicity of the metrics and our design of the CUSUM procedure with a closed-form solution for the Fréchet mean, which can be computed quickly.

5.3 Human Action Recognition

Human action recognition is another real-world application of change point detection, where a change point typically refers to abrupt changes in physical activity. We evaluate Rio-CPD using both the Log-Euclidean and Log-Cholesky metrics on the WISDM dataset [60], which contains 3-dimensional accelerometer measurements collected from a smartphone at a sampling rate of 20 Hz. The dataset includes 17 human activity changes, corresponding to 17 change points. Following prior work [45], the dataset is preprocessed through sub-sampling.

We also compare Rio-CPD against MSSA-CPD and Contra-CPD in terms of F1 score and detection delay, as previous results have consistently shown these methods outperform KL-CPD and BOCPDMS. Unless stated otherwise, the sliding window size for Rio-CPD is set to $W = 20$. The results are presented in Table 4.

We again observe that Rio-CPD, with both Riemannian metrics, significantly outperforms other methods, especially in terms of the

Table 4: F1 score and average delay (AD) on WISDM dataset.

Algorithm	F1 (Default)	F1 (Best)	AD
MSSA-CPD	0.225	0.406	15
Contra-CPD	0.314	0.378	23
Rio-CPD (**LE**)	0.378	0.453	15
Rio-CPD (**LC**)	**0.400**	**0.495**	14

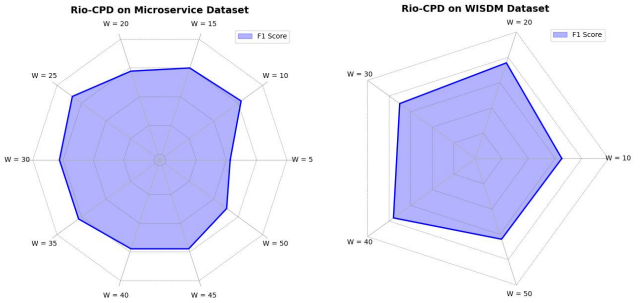

Figure 5: Sensitivity analysis of window size W for Rio-CPD.

F1 score. Moreover, the Log-Cholesky metric slightly outperforms the Log-Euclidean metric, as presented in Table 2.

5.4 Parameter Sensitivity Analysis

We outline the parameters used for Rio-CPD. Recall that Rio-CPD has one "hard" parameter—the sliding window size—and two "soft" parameters: the threshold ρ in CUSUM and L, which represents the lag between sampled data. For all datasets except HASC, we set $L = 1$, as HASC contains a long sequence, for which we use $L = 5$. The window size W is set to 10 for the Beedance dataset, 20 for Microservice and HASC, and 5 for all synthetic datasets. For the CUSUM threshold, we tested a heuristic approach by setting ρ to three times the variance of the pre-change distances, along with small integer values not greater than 5.

The performance of online algorithms is usually subject to the choice of window size. As previously discussed, we did not extensively optimize this parameter in our experiments, as our goal was to evaluate the robustness of Rio-CPD against parameter choice. To demonstrate this point, we analyze the sensitivity of W in Figure 5. We run Rio-CPD on the Microservice and WISDM datasets with different choice of W and report the F1 score results. It can be observed that with a reasonable range for W, the performance stays stable.

6 Related Work

Change Point Detection has been studied in both offline and online settings. The *offline CPD* problem is often approached as time series segmentation and has been extensively studied [14, 36, 37, 51, 52]. According to [5], change point detection methods can be categorized into four types: probabilistic, kernel-based, likelihood ratio, and subspace models. Likelihood ratio methods are typically parametric, while the others can be parametric or nonparametric. *Online CPD* has received more attention recently. The first notable

method is BOCPD (Bayesian Online Change Point Detection) [1], which inspired numerous follow-up works [2, 3, 17, 29, 38, 61]. Gaussian process-based methods [11] also fall under probabilistic models. Kernel-based approaches originated from kernel maximum mean discrepancy for change point detection [23], later refined for online settings [22, 33, 34]. Recently, neural network-based approaches [12, 13, 32] have also been classified as kernel-based methods. On the other hand, likelihood ratio methods, with the longest history, began with CUSUM [42] and the generalized likelihood ratio test [25], followed by many variants [20, 48, 57–59]. These methods are parametric, assuming known distributions. Subspace models assume that time series data lie in a low-dimensional manifold, approximated through singular spectrum analysis [4], metric design [15, 19, 28], or other learning-based approaches [21, 26, 27]. Our Rio-CPD model falls into the category of refined subspace model based methods, but has merits on accuracy and efficiency by exploring the Riemannian geometry of correlation matrices. Also different from existing work, Rio-CPD is capable of detecting both magnitude-based changes and correlation-based changes.

Riemannian Metrics on the Space of $\mathcal{S}_+^n$. The manifold $\mathcal{S}_+^n$ represents the space of positive definite matrices of dimension n, which can be equipped with various metrics. Canonical choices include the Log-Euclidean metric [6], Log-Cholesky metric [35], affine-invariant metrics [41, 43], and Bures-Wasserstein metrics [9, 18, 50]. The Log-Euclidean and Log-Cholesky metrics exhibit Lie-group bi-invariance and provide simple closed forms for the Fr'echet average of SPD matrices. The affine-invariant metric is inverse-consistent and invariant under congruence actions, such as affine transformations, and is geodesically complete. The Bures-Wasserstein metric, while not geodesically complete, is bounded by the positive semi-definite cone. Recent works also explore the Riemannian geometry of correlation matrices [16, 24], as well as more general positive semi-definite matrices [39, 53].

7 Discussion

Riemannian Geometry for Change point Detection. Geometry-inspired methods have recently gained attention in data mining and machine learning [7, 31]; however, their application to the CPD problem remains underexplored. In contrast, the Riemannian metric—a key concept in differential geometry—offers significant potential due to the broad applicability of symmetric positive definite (SPD) matrices, denoted as $\mathcal{S}_+^n$. Our framework is designed to accommodate all well-defined Riemannian metrics on $\mathcal{S}_+^n$. We have chosen the Log-Euclidean and Log-Cholesky metrics for integration into Rio-CPD because of their efficiency and simplicity. Other metrics could be incorporated using the Karcher mean, a generalized version of the Fréchet mean, computed via Riemannian optimization methods. However, this approach may involve more complex techniques and could impact the efficiency or accuracy of the Rio-CPD framework.

Dual Ability to Detect Change Points based on Magnitude and Correlation. We give insights into why Rio-CPD is capable of detecting changes on both joint and marginal distributions. We start from the intuition behind the algorithm design.

Rio-CPD integrates two complementary aspects of time series change detection.

- **Intra-relationship modeling:** Correlations between time series attributes are explicitly captured using correlation matrices. These matrices provide a structured representation of the relationships among variables, allowing the model to detect changes in these relationships.
- **Inter-relationship modeling:** Temporal dynamics are modeled by incorporating changes over time through a sliding window. This primarily tracks variations in correlations and implicitly captures changes in magnitudes within the sliding window.

For instance, even when correlations remain stable, a shift in magnitudes introduces variations in the geodesic distances between correlation matrices over successive windows. These variations are incorporated into the CUSUM statistic, enablingRio-CPD to identify magnitude-based changes alongside correlation-based ones.

A Toy Example. We construct a synthetic time series with three attributes ($\mathbf{A}, \mathbf{B}, \mathbf{C}$) and 10 time steps:

- $\mathbf{A}$: $[10, 11, 10, 9, 10, \mathbf{20}, \mathbf{20}, \mathbf{21}, \mathbf{19}, \mathbf{20}]$.
- $\mathbf{B}$: $[5, 4, 6, 4, 5, \mathbf{11}, \mathbf{10}, \mathbf{11}, \mathbf{10}, \mathbf{9}]$.
- $\mathbf{C}$: $[1, 3, 9, 5, 7, \mathbf{16}, \mathbf{11}, \mathbf{19}, \mathbf{9}, \mathbf{12}]$.

Here, $\mathbf{A}$ and $\mathbf{B}$ are correlated, while $\mathbf{C}$ is largely independent. A change point occurs at step 6, where the magnitudes of $\mathbf{A}$ and $\mathbf{B}$ double, resulting in a mean shift from ($\mathbf{A} = 10, \mathbf{B} = 5$) to ($\mathbf{A} = 20, \mathbf{B} = 10$). Using a sliding window of size 5, we compute correlation matrices for successive windows and calculate the log-Euclidean distances:

- Window $\mathbf{X}$ (steps 1–5): Correlation matrix for $[10, 11, 10, 9, 10]$, $[5, 4, 6, 4, 5]$, and $[1, 3, 9, 5, 7]$.
- Window $\mathbf{X}_1$ (steps 3–7): Correlation matrix for $[10, 9, 10, 20, 20]$, $[6, 4, 5, 11, 10]$, and $[9, 5, 7, 16, 11]$.
- Window $\mathbf{X}_2$ (steps 6–10): Correlation matrix for $[20, 20, 21, 19, 20]$, $[11, 10, 11, 10, 9]$, and $[16, 11, 19, 9, 12]$.

The geodesic distances between these matrices are:

- $d(\mathbf{X}, \mathbf{X}_1) = 5.44$
- $d(\mathbf{X}_1, \mathbf{X}_2) = 5.08$
- $d(\mathbf{X}, \mathbf{X}_2) = 2.94$

Despite the similarity in correlations between $\mathbf{X}$ and $\mathbf{X}_2$, the distances involving $\mathbf{X}_1$ are significantly larger due to the magnitude changes at step 6. With a proper parameter choice, a change point can be identified. This demonstrates Rio-CPD's ability to detect magnitude-based change points using the same geodesic distance framework applied to correlation matrices.

8 Conclusion

We address the novel problem of Correlation-aware Online Change Point Detection, which has become a crucial task in many real-world applications. We propose Rio-CPD, a non-parametric framework inspired by the Riemannian geometry of correlation matrices. We instantiate Rio-CPD with the Log-Euclidean and Log-Cholesky metrics, due to their simplicity and stability, and introduce a novel design of the CUSUM statistics. We evaluate Rio-CPD using these metrics across various datasets, both with and without correlation changes, demonstrating its superior performance in terms of accuracy and efficiency. Potential future directions include extension to high-dimensional data, and enhancement with explainability.

GenAI Usage Disclosure

We confirm that we did not use Generative AI tools in any format at any stage of this research.

References

[1] Ryan Prescott Adams and David JC MacKay. 2007. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742* (2007).

[2] Diego Agudelo-España, Sebastian Gomez-Gonzalez, Stefan Bauer, Bernhard Schölkopf, and Jan Peters. 2020. Bayesian online prediction of change points. In *Conference on Uncertainty in Artificial Intelligence*. PMLR, 320–329.

[3] Réda Alami, Odalric Maillard, and Raphael Féraud. 2020. Restarted Bayesian online change-point detector achieves optimal detection delay. In *International conference on machine learning*. PMLR, 211–221.

[4] Arwa Alanqary, Abdullah Alomar, and Devavrat Shah. 2021. Change point detection via multivariate singular spectrum analysis. *Advances in Neural Information Processing Systems* 34 (2021), 23218–23230.

[5] Samaneh Aminikhanghahi and Diane J Cook. 2017. A survey of methods for time series change point detection. *Knowledge and information systems* 51, 2 (2017), 339–367.

[6] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. 2007. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications* 29, 1 (2007), 328–347.

[7] Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. 2021. Geometric deep learning on molecular representations. *Nature Machine Intelligence* 3, 12 (2021), 1023–1032.

[8] Ian Barnett and Jukka-Pekka Onnela. 2016. Change point detection in correlation networks. *Scientific reports* 6, 1 (2016), 18893.

[9] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. 2019. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae* 37, 2 (2019), 165–191.

[10] Jedelyn Cabrieto, Francis Tuerlinckx, Peter Kuppens, Mariel Grassmann, and Eva Ceulemans. 2017. Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior research methods* 49 (2017), 988–1005.

[11] Edoardo Caldarelli, Philippe Wenk, Stefan Bauer, and Andreas Krause. 2022. Adaptive gaussian process change point detection. In *International Conference on Machine Learning*. PMLR, 2542–2571.

[12] Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. 2019. Kernel change-point detection with auxiliary deep generative models. *arXiv preprint arXiv:1901.06077* (2019).

[13] Xiuyuan Cheng and Yao Xie. 2021. Neural tangent kernel maximum mean discrepancy. *Advances in Neural Information Processing Systems* 34 (2021), 6658–6670.

[14] Fu-Lai Chung, Tak-Chung Fu, Vincent Ng, and Robert WP Luk. 2004. An evolutionary approach to pattern-based time series segmentation. *IEEE transactions on evolutionary computation* 8, 5 (2004), 471–489.

[15] AFB Costa and MA Rahim. 2006. A single EWMA chart for monitoring process mean and process variance. *Quality Technology & Quantitative Management* 3, 3 (2006), 295–305.

[16] Paul David and Weiqing Gu. 2019. A Riemannian structure for correlation matrices. *Operators and Matrices* 13, 3 (2019), 607–627.

[17] Chengyuan Deng, Zhengzhang Chen, Xujiang Zhao, Haoyu Wang, Junxiang Wang, Haifeng Chen, and Jie Gao. 2024. RIO-CPD: A Riemannian Geometric Method for Correlation-aware Online Change Point Detection. *arXiv preprint arXiv:2407.09698* (2024).

[18] Ian L Dryden, Alexey Koloydenko, and Diwei Zhou. 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. (2009).

[19] Paromita Dubey and Hans-Georg Müller. 2020. Fréchet change-point detection. *The Annals of Statistics* 48, 6 (2020), 3312–3335.

[20] Thomas Flynn and Shinjae Yoo. 2019. Change detection with the kernel cumulative sum algorithm. In *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 6092–6099.

[21] Cheng-Der Fuh. 2020. Asymptotically optimal change point detection for composite hypothesis in state space models. *IEEE Transactions on Information Theory* 67, 1 (2020), 485–505.

[22] Dian Gong, Gérard Medioni, Sikai Zhu, and Xuemei Zhao. 2012. Kernelized temporal cut for online temporal segmentation and recognition. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12*. Springer, 229–243.

[23] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. 2006. A kernel method for the two-sample-problem. *Advances in neural information processing systems* 19 (2006).

[24] Igor Grubišić and Raoul Pietersz. 2007. Efficient rank reduction of correlation matrices. *Linear algebra and its applications* 422, 2-3 (2007), 629–653.

[25] Fredrik Gustafsson. 1996. The marginalized likelihood ratio test for detecting abrupt changes. *IEEE Transactions on automatic control* 41, 1 (1996), 66–78.

[26] Yuchen Jiao, Yanxi Chen, and Yuantao Gu. 2018. Subspace change-point detection: A new model and solution. *IEEE Journal of Selected Topics in Signal Processing* 12, 6 (2018), 1224–1239.

[27] Yoshinobu Kawahara, Takehisa Yairi, and Kazuo Machida. 2007. Change-point detection in time-series data based on subspace identification. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, 559–564.

[28] Nicolas Keriven, Damien Garreau, and Iacopo Poli. 2020. NEWMA: a new method for scalable model-free online change-point detection. *IEEE Transactions on Signal Processing* 68 (2020), 3515–3528.

[29] Jeremias Knoblauch and Theodoros Damoulas. 2018. Spatio-temporal Bayesian on-line changepoint detection with model selection. In *International Conference on Machine Learning*. PMLR, 2718–2727.

[30] Marc Lavielle. 1999. Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their applications* 83, 1 (1999), 79–102.

[31] Brigitte Le Roux and Henry Rouanet. 2004. *Geometric data analysis: from correspondence analysis to structured data analysis*. Springer Science & Business Media.

[32] Jie Li, Paul Fearnhead, Piotr Fryzlewicz, and Tengyao Wang. 2024. Automatic change-point detection in time series via deep learning. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86, 2 (2024), 273–285.

[33] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. 2015. M-statistic for kernel change-point detection. *Advances in Neural Information Processing Systems* 28 (2015).

[34] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. 2019. Scan B-statistic for kernel change-point detection. *Sequential Analysis* 38, 4 (2019), 503–544.

[35] Zhenhua Lin. 2019. Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* 40, 4 (2019), 1353–1370.

[36] Xiaoyan Liu, Zhenjiang Lin, and Huaiqing Wang. 2008. Novel online methods for time series segmentation. *IEEE Transactions on knowledge and data engineering* 20, 12 (2008), 1616–1626.

[37] Miodrag Lovrić, Marina Milanović, and Milan Stamenković. 2014. Algoritmic methods for segmentation of time series: An overview. *Journal of Contemporary Economic and Business Issues* 1, 1 (2014), 31–53.

[38] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. 2013. Online Bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar conference on signals, systems and computers*. IEEE, 1833–1837.

[39] Estelle Massart and P-A Absil. 2020. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM J. Matrix Anal. Appl.* 41, 1 (2020), 171–198.

[40] Iris B Mauss, Robert W Levenson, Loren McCarter, Frank H Wilhelm, and James J Gross. 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 2 (2005), 175.

[41] Maher Moakher. 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications* 26, 3 (2005), 735–747.

[42] Ewan S Page. 1954. Continuous inspection schemes. *Biometrika* 41, 1/2 (1954), 100–115.

[43] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. 2006. A Riemannian framework for tensor computing. *International Journal of computer vision* 66 (2006), 41–66.

[44] Andrey Pepelyshev and Aleksey S Polunchenko. 2015. Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570* (2015).

[45] Nikita Puchkin and Valeriia Shcherbakova. 2023. A contrastive approach to online change point detection. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 5686–5713.

[46] Friedrich Pukelsheim. 1994. The three sigma rule. *The American Statistician* 48, 2 (1994), 88–91.

[47] Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology* 46, 6 (2007), 900–915.

[48] Gaetano Romano, Idris A Eckley, Paul Fearnhead, and Guillem Rigaill. 2023. Fast online changepoint detection via functional pruning CUSUM statistics. *Journal of Machine Learning Research* 24, 81 (2023), 1–36.

[49] David Siegmund. 2013. *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

[50] Asuka Takatsu. 2011. Wasserstein geometry of Gaussian measures. (2011).

[51] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2020. Selective review of offline change point detection methods. *Signal Processing* 167 (2020), 107299.

[52] Gerrit JJ Van den Burg and Christopher KI Williams. 2020. An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222* (2020).

[53] Bart Vandereycken, P-A Absil, and Stefan Vandewalle. 2013. A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA J. Numer. Anal.* 33, 2 (2013), 481–514.

[54] Dongjie Wang, Zhengzhang Chen, Yanjie Fu, Yanchi Liu, and Haifeng Chen. 2023. Incremental causal graph learning for online root cause analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

2269–2278.

[55] Dongjie Wang, Zhengzhang Chen, Jingchao Ni, Liang Tong, Zheng Wang, Yanjie Fu, and Haifeng Chen. 2023. Hierarchical graph neural networks for causal discovery and root cause localization. *arXiv preprint arXiv:2302.01987* (2023).

[56] Dongjie Wang, Zhengzhang Chen, Jingchao Ni, Liang Tong, Zheng Wang, Yanjie Fu, and Haifeng Chen. 2023. Interdependent causal networks for root cause localization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5051–5060.

[57] Haoyun Wang and Yao Xie. 2024. Sequential change-point detection: Computation versus statistical performance. *Wiley Interdisciplinary Reviews: Computational Statistics* 16, 1 (2024), e1628.

[58] Kes Ward, Giuseppe Dilillo, Idris Eckley, and Paul Fearnhead. 2023. Poisson-FOCuS: An efficient online method for detecting count bursts with application to gamma ray burst detection. *J. Amer. Statist. Assoc.* (2023), 1–13.

[59] Song Wei and Yao Xie. 2022. Online kernel CUSUM for change-point detection. *arXiv preprint arXiv:2211.15070* (2022).

[60] Gary M Weiss, Kenichi Yoneda, and Thaier Hayajneh. 2019. Smartphone and smartwatch-based biometrics using activities of daily living. *Ieee Access* 7 (2019), 133190–133202.

[61] Xiang Xuan and Kevin Murphy. 2007. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th international conference on Machine learning*. 1055–1062.

[62] Ping Yang, Guy Dumont, and John Mark Ansermino. 2006. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on biomedical engineering* 53, 11 (2006), 2211–2219.

[63] Jennifer Yu, Tina Behrouzi, Kopal Garg, Anna Goldenberg, and Sana Tonekaboni. 2023. Dynamic Interpretable Change Point Detection for Physiological Data Analysis. In *Machine Learning for Health (ML4H)*. PMLR, 636–649.

[64] Ruohong Zhang, Yu Hao, Donghan Yu, Wei-Cheng Chang, Guokun Lai, and Yiming Yang. 2020. Correlation-aware change-point detection via graph neural networks. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23–27, 2020, Proceedings, Part III 27*. Springer, 555–567.

[65] Lecheng Zheng, Zhengzhang Chen, Haifeng Chen, and Jingrui He. 2024. Online multi-modal root cause analysis. *arXiv preprint arXiv:2410.10021* (2024).

[66] Lecheng Zheng, Zhengzhang Chen, Jingrui He, and Haifeng Chen. 2024. MULAN: multi-modal causal structure learning and root cause analysis for microservice systems. In *Proceedings of the ACM Web Conference 2024*. 4107–4116.

[67] Lecheng Zheng, Zhengzhang Chen, Dongjie Wang, Chengyuan Deng, Reon Matsuoka, and Haifeng Chen. 2024. LEMMA-RCA: A Large Multi-modal Multi-domain Dataset for Root Cause Analysis. *arXiv preprint arXiv:2406.05375* (2024).