RIO-CPD: A Riemannian Geometric Method for Correlation-aware Online Change Point Detection

Chengyuan Deng [1] **Zhengzhang Chen** [2] **Xujiang Zhao** [2] **Haoyu Wang** [2]
Junxiang Wang [2] **Haifeng Chen** [2] **Jie Gao** [1]

Abstract

The objective of change point detection is to identify abrupt changes at potentially multiple points within a data sequence. This task is particularly challenging in the online setting where various types of changes can occur, including shifts in both the marginal and joint distributions of the data. This paper tackles these challenges by sequentially tracking correlation matrices on the Riemannian geometry, where the geodesic distances accurately capture the development of correlations. We propose RIO-CPD, a non-parametric correlation-aware online change point detection framework that combines the Riemannian geometry of the manifold of symmetric positive definite matrices and the cumulative sum statistic (CUSUM) for detecting change points. RIO-CPD enhances CUSUM by computing the geodesic distance from present observations to the Fréchet mean of previous observations. With careful choice of metrics equipped to the Riemannian geometry, RIO-CPD is simple and computationally efficient. Experimental results on both synthetic and real-world datasets demonstrate that RIO-CPD outperforms existing methods in detection accuracy and efficiency.

1. Introduction

The task of change point detection (Page, 1954) seeks to identify abrupt distributional changes in the temporal evolution of a system through noisy observations. Localizing such changes not only helps in isolating and interpreting different patterns of the time series data but also has crucial consequences for the safety and reliability of the system. The change point detection (CPD) problem has found applications in various domains including climatology (Reeves et al., 2007; Gallagher et al., 2013), finance (Pepelyshev & Polunchenko, 2015; Lavielle & Teyssiere, 2007), healthcare (Yang et al., 2006), among others. We focus on unsupervised *online* change point detection in the discrete setting where the observations are streamed as multivariate time series, and multiple change points may emerge.

Although recent years have witnessed significant progress in online CPD techniques (Keriven et al., 2020; Alanqary et al., 2021; Caldarelli et al., 2022; Li et al., 2024; Wu et al., 2023), it remains a challenge to efficiently capture changes in various patterns, such as those in the marginal distribution (*e.g.*, independent magnitude) and joint distribution (*e.g.*, correlation between covariates). The capability of correlation-awareness in CPD (Lavielle, 1999; Barnett & Onnela, 2016; Cabrieto et al., 2018; 2017; Zhang et al., 2020; Yu et al., 2023) has recently garnered increased attention and has had a practical impact on fields including behavioral science (Mauss et al., 2005) and root cause analysis in AIOps (Wang et al., 2023a;d; Chen et al., 2024). However, these methods are not designed for online use. This is because extracting correlations may involve complex techniques such as Graph Neural Networks or probabilistic graphical models, which can be inherently difficult to adapt to an online setting. Moreover, these approaches often require extended processing times. Therefore, there is a vital need for efficient and accurate correlation-aware online CPD methods capable of detecting multiple change points.

To fill this gap, we propose RIO-CPD, a non-parametric Riemannian metrics based method for correlation-aware Online Change Point Detection. RIO-CPD applies Riemannian metrics, such as Log-Euclidean or Log-Cholesky, on a Riemannian manifold of correlation matrices. These metrics allow for the accurate computation of geodesic distances and the Fréchet mean from a set of correlation matrices, which in turn are utilized to construct the cumulative sum (CUSUM) statistic. This statistic performs sequential hypothesis testing to determine whether a time step is a change point. The method's natural correlation-awareness stems from corre-

Work done during an internship at NEC Laboratories America. [1]Rutgers University [2]NEC Laboratories America. Correspondence to: Zhengzhang Chen <zchen@nec-labs.com>.

Proceedings of the Geometry-grounded Representation Learning and Generative Modeling at the 41st International Conference on Machine Learning, Vienna, Austria. PMLR Vol Number, 2024.

lation matrices where the entries are Pearson correlation coefficients. RIO-CPD extends the overarching framework of CPD algorithms known as the *subspace model* (Kawahara et al., 2007; Aminikhanghahi & Cook, 2017), providing several merits. An overview comparing RIO-CPD to methods based on the subspace model is illustrated in Figure 1.

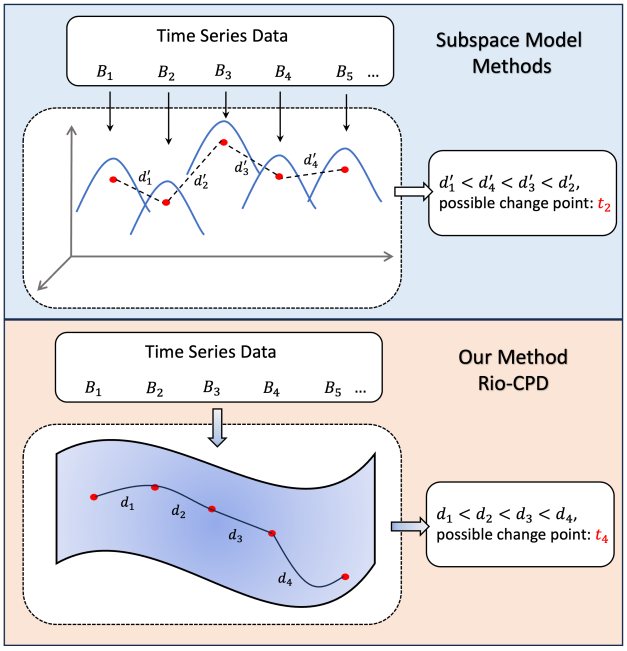

Figure 1. Framework of RIO-CPD and subspace model methods. B_i indicates the i-th batch of time series data, and d_i denotes the distances between neighboring batches on the subspace.

For existing works based on the subspace model, a low-dimensional subspace is first estimated for data streamed through a sliding window, using various techniques such as dimension reduction (Jiao et al., 2018) and singular spectrum analysis (Alanqary et al., 2021). Next, *Euclidean* metrics are applied to the subspace to measure distances between consecutive time steps, where abruptly different distances imply a higher probability of a change point. However, this framework has limitations due to distortions from both the estimated subspace, which can misrepresent the original data, and the assumption that the subspace adheres to Euclidean geometry, which it may not. RIO-CPD addresses these limitations by (1) applying directly to the Riemannian manifold of correlation matrices, rather than an estimated subspace; and (2) using a *Riemannian* metric, which more accurately captures the manifold's geometry. This approach allows for more precise and efficient measurement of correlation developments, as it bypasses the subspace learning process.

Geometry-inspired methods have been actively studied recently in data mining and machine learning (Le Roux & Rouanet, 2004; Atz et al., 2021; Kirby, 2000); however, they are still under-explored for the CPD problem. On the other hand, the Riemannian metric, a fundamental topic in differential geometry, is especially motivated by the diverse applications of symmetric positive definite (SPD) matrices $\mathcal{S}_+^n$. Our framework, in general, accommodates all well-defined Riemannian metrics on $\mathcal{S}_+^n$. We have selected the Log-Euclidean and Log-Cholesky metrics for integration into RIO-CPD due to their efficiency and simplicity, as discussed in Section 4.4.

Our contributions are summarized as follows:

- We propose RIO-CPD, a framework for online change point detection that utilizes the Riemannian geometry of correlation matrices and the CUSUM procedure. RIO-CPD is capable of detecting changes both independently and in correlation, and it is both non-parametric and highly efficient.
- The proposed framework has been extensively evaluated on both synthetic and real-world datasets. The results demonstrate that RIO-CPD significantly outperforms state-of-the-art methods in terms of detection accuracy and efficiency.

2. Related Work

Online Change Point Detection. There has been a rich literature of change point detection in both offline and online fashion, while the latter gathers more attention in recent years. Following the taxonomy outlined by (Aminikhanghahi & Cook, 2017), we consider methods for online change point detection based on four models: Probabilistic Model, Subspace Model, Kernel-based Model and Likelihood Ratio Model. The earliest attempt is BOCPD (Bayesian online change point detection) (Adams & MacKay, 2007) and many follow-up works (Xuan & Murphy, 2007; Malladi et al., 2013; Agudelo-España et al., 2020; Alami et al., 2020; Knoblauch & Damoulas, 2018) using Bayesian methods. Gaussian process-based methods (Caldarelli et al., 2022) also fall into this family. On the other hand, subspace model assumes the time series data lies in a low-dimensional manifold, which can be approximated using techniques such as singular spectrum analysis (Alanqary et al., 2021) or other learning methods (Jiao et al., 2018; Kawahara et al., 2007). Recently, a number of approaches employing neural networks have been proposed (Li et al., 2024), and can be regarded as kernel-based methods. As for likelihood ratio methods, CUSUM (Page, 1954) and its variants (Flynn & Yoo, 2019; Romano et al., 2023) are widely used. Due to the large load of research on this topic, we refer to survey papers (Aminikhanghahi & Cook, 2017; Van den Burg & Williams, 2020) for further interest.

Riemannian metrics on the space of $\mathcal{S}_+(n)$. The manifold of $\mathcal{S}_+(n)$ can be equipped with a variety of metrics. The canonical choices include Log-Euclidean metric (Ar-

signy et al., 2007), Log-Cholesky metric (Lin, 2019), affine-invariant metrics (Moakher, 2005; Pennec et al., 2006) and Bures-Wasserstein metrics (Dryden et al., 2009; Takatsu, 2011; Bhatia et al., 2019). The Log-Euclidean and Log-Cholesky metric exhibit Lie-group bi-invariance and have simple closed form of the Fréchet average of SPD matrices. The affine-invariant metric is inverse-consistent and invariant under congruence actions such as affine transformations. They are all geodesically complete. The Bures-Wasserstien metric, however, is not geodesically complete but is bounded by the positive semi-definite cone. Recent works also study the Riemannian geometry of correlation matrices (Grubišić & Pietersz, 2007; David & Gu, 2019), also more generally positive semi-definite matrices (Vandereycken et al., 2013; Massart & Absil, 2020).

3. 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 $t \in [T]$. Further, 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 problem 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 on 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, namely, we aim for $\hat{\tau} > \tau$ and $\hat{\tau} - \tau$ is small.

4. Methodology

The starting point of our algorithm is the fact the correlation matrices, in practice[1], lie in $\mathcal{S}_+^n$ that exhibit Riemannian structures (Arsigny et al., 2007). This allows for a Riemannian metric to be applied to correlation matrices with low distortions. More specifically, we calculate the Fréchet mean of the set of correlation matrices in previous observations, next the distance between the Fréchet mean and the present correlation matrix. We construct the CUSUM statistic (Section 4.1) based on this step and perform sequential hypothesis testing for change point detection. We now introduce two Riemannian metrics, known as Log-Euclidean and Log-Cholesky metric, together with their closed form Fréchet mean. The closed form is non-trivial and circumvents RIO-CPD from Riemannian optimization. We wrap up this section by presenting the algorithm in Section 4.4.

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

4.1. CUSUM Procedure

The CUSUM statistic is measure of the possibility of a timestamp being a change point, and has been widely used in the manner of hypothesis testing. To construct the CUSUM statistic at timestamp t, a detection score is required, which usually evaluates the "distance" between the observation at timestamp t and the distribution of the time series streamed so far. Formally, we have:

Definition 4.1 (CUSUM Statistic)**.** Given an observation $X(t)$ at time step t and and a detection score $D(t)$, the CUSUM statistic of time series $\{X(1), \ldots, X(t)\}$ is $y(t) = \max_{1 \leq i \leq t} \sum_{j=i}^{t} D(j)$.

It is well-known from the literature that Definition 4.1 can be alternatively written as $y(t) = \max\{y(t-1) + D(t), 0\}$, where $y(0) = 0$. The recursive computation improves the efficiency of testing based on only the history one step before. The CUSUM procedure thus proceeds as follows: At each timestamp t, perform the hypothesis testing

$$H(t) = H_{\mathbb{1}_{y(t) \geq \tau}}, \text{ where } H_0 : t \text{ is a not change point and } H_1 : t \text{ is a change point.}$$

Note that $\rho > 0$ is a threshold. The set of change points is therefore $\{t | y(t) \geq \rho\}$.

At the first sight, the CUSUM procedure provides a simple framework for change point detection, however, the design of the detection score $D(t)$ can be non-trivial and usually determines the performance. We explain how our method constructs $D(t)$ in the following sections. An essential property of $D(t)$ requires a separation between change points and regular ones, that is $\mathbb{E}[D(t)|H_0] < 0$ and $\mathbb{E}[D(t)|H_1] > 0$.

4.2. Log-Euclidean Metric and Fréchet Mean

We first introduce the Lie group structure of $\mathcal{S}_+^n$, which empowers a bi-invariant Riemannian metric induced by 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 (Definition A.2).

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

where λ is a real number. 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:

$$d_{LE}(P_1, P_2) = \|\phi_{LE}(P_1) - \phi_{LE}(P_2)\|_F \quad (1)$$

The computation of Log-Euclidean metric is demanding theoretically due to the computation of logarithm map, however, there are techniques that reduce the computation time. Another nice property of the Log-Euclidean metric is it gives a closed form of the Fréchet mean of SPD matrices, which can be regarded as the "centroid" of $\mathcal{S}_+^n$ geometry. Similar as the centroid of Euclidean space (*e.g.*, k-means objective), the Fréchet mean minimizes the square error of the geodesic distances to all SPD matrices (Definition A.3).

The Fréchet mean of Log-Euclidean geometry (also known as Log-Euclidean mean) is given by (Arsigny et al., 2007):

$$\sigma_{LE}(P_1, \ldots, P_n) = \exp\big(\frac{1}{n}\sum_{i=1}^{n}\phi_{LE}(P_i)\big) \quad (2)$$

Note that the Log-Euclidean mean is a natural generalization of the geometric mean. If P_i are positive real numbers, their geometric mean follow exactly the same formula. Log-Euclidean and Log-Cholesky are the only two Riemannian metrics that admit a closed form of the Fréchet mean. For other metrics, the Fréchet mean may not be even unique.

4.3. Log-Cholesky Metric and Fréchet Mean

The Log-Euclidean and Log-Cholesky metric (Lin, 2019) share a number of similarities, and at a high level, are topologically same because $\mathcal{L}^n \simeq \mathcal{S}^n \simeq \mathbb{R}^{n(n+1)/2}$. The Log-Euclidean is obtained from the matrix logarithm: $\mathcal{S}_+^n \to \mathcal{S}^n$ while the other is based on $\mathcal{S}_+^n \to \mathcal{L}^n$. The definition of Log-Cholesky metric starts from Cholesky decomposition, which provides a unique representation for any Hermitian and positive definite matrix A in the form of LL^T. Here L is a lower triangular matrix and L^T is upper triangular. The matrix logarithm for 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 of L with zero entries elsewhere. The geodesic distance under Log-Cholesky metric is:

$$\begin{aligned} d_{LC}(P_1, P_2) &= \|\phi_{LE}(P_1) - \phi_{LE}(P_2)\|_F \\ &= \Big(\|\lfloor L_1 \rfloor - \lfloor L_2 \rfloor\|_F^2 \\ &\quad + \|\ln\big(\mathbb{D}(L_1)\big) - \ln\big(\mathbb{D}(L_2)\big)\|_F^2\Big)^{1/2} \end{aligned} \quad (3)$$

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

The Log-Cholesky metric also enjoys a nice closed form of 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 $\mathcal{L}_+^n$:

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

4.4. RIO-CPD Algorithm

RIO-CPD only requires one parameter pre-specified: the size of the sliding window W. The threshold ρ used in CUSUM can be regarded as a parameter but can be handled by our algorithm. Other heuristics on selecting ρ will be discussed later. Our algorithm proceeds with three steps: (1) Construct correlation matrices $\mathbf{B}_t$ for each time step t; (2) Monitor the geodesic distance between $\mathbf{B}_t$ and σ_{t-1}, the Fréchet mean of previous correlation matrices. (3) Construct CUSUM statistic using the distance and perform sequential hypothesis testing. We now explain each step in details.

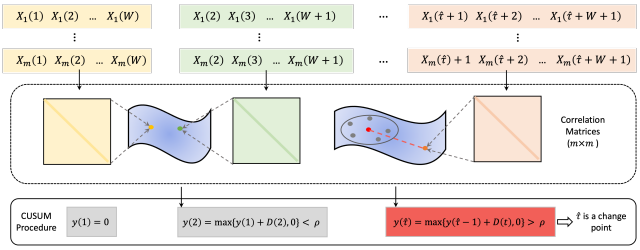

Figure 2. The overview of the proposed framework RIO-CPD.

Step 1. Transform observations in a sliding window to 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 Log-Euclidean or Log-Cholesky metric, we compute $\sigma_{t-1}^{\mathfrak{g}}$, the Fréchet mean of $\mathbf{B}_1, \ldots, \mathbf{B}_{t-1}$ and the distance from $\mathbf{B}_t$ to $\sigma_{t-1}^{\mathfrak{g}}$, *i.e.*, $d_t = \mathfrak{g}(\mathbf{B}_t, \sigma_{t-1}^{\mathfrak{g}})$. [2]

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

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

With the detection score $D(t)$, the CUSUM test repeats on computing $y(t)$ for $t > W$ and $y(W) = 0$, until a change point is detected, *i.e.*, $y(t) > \rho$ for a threshold ρ. Namely,

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

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

The procedure above can be adapted into the multiple change points setting by restarting at $\hat{\tau}+1$ after a change point $\hat{\tau}$ is detected, with a new base correlation matrix $\mathbf{B}_{\hat{\tau}+1}$. The process of our algorithm is illustrated in Figure 2. More intuition and discussions on our algorithm design are presented in Appendix B.

5. Experiments

5.1. Datasets and Baselines

The *synthetic data* is based on particle-spring system adopted from (Zhang et al., 2020) and contains correlation changes. For *real-world data*, we select commonly-compared benchmark datasets including Beedance[3] and HASC[4], which do not necessarily have change points based on correlations. Further, we include a Product Review Microservice dataset (Zheng et al., 2024c;a;b; Wang et al., 2023c;b) with correlation changes. The basic statistics are presented in Table 1, more details on datasets can be found in Appendix C.1.

Table 1. Datasets used in experiments.

Dataset	Length	# Change Points	# Time Series	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	$\mathbb{R}^3$	✗

We compare with recent SOTA baselines, encompassing different families of online CPD methods: Kernel-based method (KL-CPD (Chang et al., 2019)), Probabilistic Model method (Spatial-temporal Bayesian Online CPD (Knoblauch & Damoulas, 2018)), and Subspace Model method (Multivariate Singular Spectrum Analysis (Alanqary et al., 2021)).

5.2. Performance

We evaluate the performance of RIO-CPD on accuracy by the F1-score, and report results in two settings: *Default* indicates vanilla parameter initialization and *Best* has a fine-tuning process. We count a change point successfully detected if it lies in the sliding window reported by the algorithm. The results on synthetic datasets and real-world datasets are demonstrated in Table 2. We also report the average delay of detection, which is a crucial measure for online CPD and the running time in Table 3. To separate from false alarm, average delay is counted for points identified within twice window-size delay.

We conclude from experiment results that RIO-CPD with both Log-Euclidean (LE) and Log-Cholesky (LC) metrics consistently performs better than other methods or at least competitively. Especially for datasets with correlation-based change points, RIO-CPD gives better results by a large margin. We observe from the experiments that other methods suffer from both false negative and false positive outputs: they tend to miss correlation changes on one hand and report more change points due to local perturbation on the other hand. KL-CPD on Microservice data has 0 F1-score because there is no true positive detection. Another remark is that the LC metric is claimed to be more numerically stable and computationally efficient (Lin, 2019) than the LE metric, as shown by our experiments, RIO-CPD with LC metric indeed slightly performs better. We provide more details on the implementation in Appendix C.2.

Table 2. Detection Performance on Real and Synthetic Datasets.

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
Multivariate SSA	0.154	0.308	0.500	**0.659**	0.177	0.327
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
Multivariate SSA	0.179	0.330	0.292	0.485	0.153	0.308
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

Table 3. Average Delay and Computation Time on Real-world and Synthetic Datasets. We put N.A. for Average Delay only if no positive detection falls into a window of reasonable delay.

Algorithm	Microservice		Beedance		HASC	
	Ave Delay	Time	Ave Delay	Time	Ave Delay	Time
KL-CPD	N.A.	>5min	7	> 5min	N.A.	>15min
BOCPDMS	N.A.	100s	N.A.	35s	N.A.	10min
Multivariate SSA	14	0.5s	11	0.5s	41	12s
RIO-CPD (**LE**)	0	0.7s	8	0.5s	25	16s
RIO-CPD (**LC**)	0	0.5s	6	0.5s	19	13s
	Connection		Speed		Location	
	Ave Delay	Time	Ave Delay	Time	Ave Delay	Time
KL-CPD	N.A.	>15min	N.A.	>15min	N.A.	>15min
BOCPDMS	N.A.	4min	N.A.	4min	N.A.	4min
Multivariate SSA	2	7s	4	7s	N.A.	7s
RIO-CPD (**LE**)	2	8s	2	8s	2	8s
RIO-CPD (**LC**)	2	7s	2	7s	3	7s

6. Conclusion

We propose RIO-CPD, a correlation-aware online CPD method inspired by the Riemannian geometric structure of correlation matrices. RIO-CPD serves as a general framework that employs Riemannian metrics, particularly benefiting from the simplicity and stability of the Log-Euclidean and Log-Cholesky metrics. We have evaluated RIO-CPD using these metrics across various datasets, both with and without correlation changes, demonstrating its superior performance in terms of accuracy and efficiency.

[3] https://sites.cc.gatech.edu/ borg/ijcv_psslds/

[4] http://hasc.jp/hc2011/

References

Adams, R. P. and MacKay, D. J. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

Agudelo-España, D., Gomez-Gonzalez, S., Bauer, S., Schölkopf, B., and Peters, J. Bayesian online prediction of change points. In *Conference on Uncertainty in Artificial Intelligence*, pp. 320–329. PMLR, 2020.

Alami, R., Maillard, O., and Féraud, R. Restarted bayesian online change-point detector achieves optimal detection delay. In *International conference on machine learning*, pp. 211–221. PMLR, 2020.

Alanqary, A., Alomar, A., and Shah, D. Change point detection via multivariate singular spectrum analysis. *Advances in Neural Information Processing Systems*, 34: 23218–23230, 2021.

Aminikhanghahi, S. and Cook, D. J. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1):328–347, 2007.

Atz, K., Grisoni, F., and Schneider, G. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 3(12):1023–1032, 2021.

Barnett, I. and Onnela, J.-P. Change point detection in correlation networks. *Scientific reports*, 6(1):18893, 2016.

Bhatia, R., Jain, T., and Lim, Y. On the bures–wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019.

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., and Ceulemans, E. Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior research methods*, 49:988–1005, 2017.

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., and Ceulemans, E. Testing for the presence of correlation changes in a multivariate time series: A permutation based approach. *Scientific reports*, 8(1):769, 2018.

Caldarelli, E., Wenk, P., Bauer, S., and Krause, A. Adaptive gaussian process change point detection. In *International Conference on Machine Learning*, pp. 2542–2571. PMLR, 2022.

Chang, W.-C., Li, C.-L., Yang, Y., and Póczos, B. Kernel change-point detection with auxiliary deep generative models. *arXiv preprint arXiv:1901.06077*, 2019.

Chen, Z., Chen, H., Tong, L., and Wang, D. Trigger point detection for online root cause analysis and system fault diagnosis, February 15 2024. US Patent App. 18/359,288.

David, P. and Gu, W. A riemannian structure for correlation matrices. *Operators and Matrices*, 13(3):607–627, 2019.

Dryden, I. L., Koloydenko, A., and Zhou, D. Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. 2009.

Flynn, T. and Yoo, S. Change detection with the kernel cumulative sum algorithm. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 6092–6099. IEEE, 2019.

Gallagher, C., Lund, R., and Robbins, M. Changepoint detection in climate time series with long-term trends. *Journal of Climate*, 26(14):4994–5006, 2013.

Grubišić, I. and Pietersz, R. Efficient rank reduction of correlation matrices. *Linear algebra and its applications*, 422(2-3):629–653, 2007.

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

Kawahara, Y., Yairi, T., and Machida, K. Change-point detection in time-series data based on subspace identification. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pp. 559–564. IEEE, 2007.

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

Kirby, M. *Geometric data analysis: an empirical approach to dimensionality reduction and the study of patterns*. John Wiley & Sons, Inc., 2000.

Knoblauch, J. and Damoulas, T. Spatio-temporal bayesian on-line changepoint detection with model selection. In *International Conference on Machine Learning*, pp. 2718–2727. PMLR, 2018.

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

Lavielle, M. and Teyssiere, G. Adaptive detection of multiple change-points in asset price volatility. In *Long memory in economics*, pp. 129–156. Springer, 2007.

Le Roux, B. and Rouanet, H. *Geometric data analysis: from correspondence analysis to structured data analysis*. Springer Science & Business Media, 2004.

Li, J., Fearnhead, P., Fryzlewicz, P., and Wang, T. Automatic change-point detection in time series via deep learning. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2):273–285, 2024.

Lin, Z. Riemannian geometry of symmetric positive definite matrices via cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications*, 40(4):1353–1370, 2019.

Malladi, R., Kalamangalam, G. P., and Aazhang, B. Online bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar conference on signals, systems and computers*, pp. 1833–1837. IEEE, 2013.

Massart, E. and Absil, P.-A. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM Journal on Matrix Analysis and Applications*, 41(1):171–198, 2020.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., and Gross, J. J. The tie that binds? coherence among emotion experience, behavior, and physiology. *Emotion*, 5(2):175, 2005.

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

Page, E. S. Continuous inspection schemes. *Biometrika*, 41 (1/2):100–115, 1954.

Pennec, X., Fillard, P., and Ayache, N. A riemannian framework for tensor computing. *International Journal of computer vision*, 66:41–66, 2006.

Pepelyshev, A. and Polunchenko, A. S. Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570*, 2015.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology*, 46(6):900–915, 2007.

Romano, G., Eckley, I. A., Fearnhead, P., and Rigaill, G. Fast online changepoint detection via functional pruning cusum statistics. *Journal of Machine Learning Research*, 24(81):1–36, 2023.

Takatsu, A. Wasserstein geometry of gaussian measures. 2011.

Van den Burg, G. J. and Williams, C. K. An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222*, 2020.

Vandereycken, B., Absil, P.-A., and Vandewalle, S. A riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA Journal of Numerical Analysis*, 33(2):481–514, 2013.

Wang, D., Chen, Z., Fu, Y., Liu, Y., and Chen, H. Incremental causal graph learning for online root cause analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2269–2278, 2023a.

Wang, D., Chen, Z., Ni, J., Tong, L., Wang, Z., Fu, Y., and Chen, H. Hierarchical graph neural networks for causal discovery and root cause localization. *arXiv preprint arXiv:2302.01987*, 2023b.

Wang, D., Chen, Z., Ni, J., Tong, L., Wang, Z., Fu, Y., and Chen, H. Interdependent causal networks for root cause localization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5051–5060, 2023c.

Wang, X., Borsoi, R. A., and Richard, C. Online change point detection on riemannian manifolds with karcher mean estimates. In *2023 31st European Signal Processing Conference (EUSIPCO)*, pp. 2033–2037. IEEE, 2023d.

Wu, S., Diao, E., Banerjee, T., Ding, J., and Tarokh, V. Score-based quickest change detection for unnormalized models. In *AISTATS*, pp. 10546–10565, 2023.

Xuan, X. and Murphy, K. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th international conference on Machine learning*, pp. 1055–1062, 2007.

Yang, P., Dumont, G., and Ansermino, J. M. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on biomedical engineering*, 53(11):2211–2219, 2006.

Yu, J., Behrouzi, T., Garg, K., Goldenberg, A., and Tonekaboni, S. Dynamic interpretable change point detection for physiological data analysis. In *Machine Learning for Health (ML4H)*, pp. 636–649. PMLR, 2023.

Zhang, R., Hao, Y., Yu, D., Chang, W.-C., Lai, G., and Yang, Y. 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*, pp. 555–567. Springer, 2020.

Zheng, L., Chen, Z., He, J., and Chen, H. Mulan: Multi-modal causal structure learning and root cause analysis for microservice systems. In *Proceedings of the ACM on Web Conference 2024*, pp. 4107–4116, 2024a.

Zheng, L., Chen, Z., He, J., and Chen, H. Multi-modal causal structure learning and root cause analysis. *arXiv preprint arXiv:2402.02357*, 2024b.

Zheng, L., Chen, Z., Wang, D., Deng, C., Matsuoka, R., and Chen, H. Lemma-rca: A large multi-modal multi-domain dataset for root cause analysis. *arXiv preprint arXiv:2406.05375*, 2024c.

A. Standard Notions in Differential Geometry

First, a manifold $\mathcal{M}$ is a topological space locally diffeomorphic to Euclidean space. Now we define the Riemannian manifold, which can be considered as a manifold coupled with a Riemannian metric.

Definition A.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 through the inner product. More precisely, for two points $p, q \in \mathcal{M}$, the geodesic distance between p, q is: $d_{\mathcal{M}}(p, q) = \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))}dt$, where $\gamma : [0, 1] \to \mathcal{M}$ and we have $\gamma(0) = p, \gamma(1) = q$.

Another fundamental notion is the exponential and logarithm map, which is the geometric connection between a Lie group and its Lie algebra, because the Lie algebra can be viewed as the tangent space to the Lie group at the identity. In our case, it builds an one-to-one correspondence between SPD matrices and the vector space structure, such that geodesics are well-defined and computable.

Definition A.2 (Exponential and Logarithm Map of Matrices)**.** Given a matrix A, its exponential map is defined as $\exp(A) := \sum_{k=0}^{\infty} A^k/k!$ and logarithm map is the inverse, denoted as $\log(A)$.

Last, we introduce the Fréchet mean of a set of matrices in a certain metric space.

Definition A.3 (Fréchet mean)**.** Let $(\mathcal{M}, d)$ be a complete metric space, $P_1, \dots P_n$ are points in $\mathcal{M}$, then the Fréchet mean of $P_1, \dots P_n$ is:

$$\sigma_{\mathcal{M}} = \operatorname*{argmin}_{x \in \mathcal{M}} \sum_{i=1}^{n} d^2(x, P_i) \tag{4}$$

B. More Discussions on Algorithm Design

As mentioned in Section 1, our algorithm is inspired by methods based on subspace model but directly applies to the Riemannian geometry of the space of $\mathcal{S}_+^n$. The distance between correlation matrices implies the possibility of a change point emergence. In Step 2 and 3, RIO-CPD aims at formalize this intuition into an algorithm that works with CUSUM seamlessly. In fact, the design of the detection score $D(t)$ is inspired by clustering, which also shares strong geometric insights. Basically, by our design, RIO-CPD is supposed to output t is a change point if $\mathbf{B}_t$ is an outlier from the cluster by $\mathbf{B}_1, \dots, \mathbf{B}_{t-1}$. To capture an outlier, we obtain the difference between the current distance and the "radius" of the cluster, if the difference is higher than a threshold, it is likely that $\mathbf{B}_t$ is an outlier, thus a change point.

As for the computation time, the bottleneck of running time for our algorithm lies in the eigen decomposition (Log-euclidean) and the Cholesky decomposition (Log-Cholesky), both are implemented in practice with $O(m^3)$ time for m-dimension symmetric matrices. Theoretically the exponent can be reduced to some constant close to ω (fast matrix multiplication constant). Note that m is the number of time series in our setting, and is usually a constant. Further, our algorithm does not involve Riemannian optimization or other learning methods. Therefore RIO-CPD is fairly efficient, as we show in Table 3.

Finally, it is natural to ask if RIO-CPD can be extended to more Riemannian metrics and even other manifolds. A crucial issue here is the calculation of the centroid (or other similar notions, such as Karcher mean). Note that the closed form solution of Fréchet mean of Log-Euclidean and Log-Cholesky geometry provides a simple way of obtaining the centroid of the $\mathcal{S}_+^n$, therefore allows for obtaining the radius. To the best of our knowledge, these are the only known Riemannian metrics for $\mathcal{S}_+^n$ that have such a nice property. Generalizing RIO-CPD to other metrics or manifold would inevitably involve Riemannnian optimization, which may work well on the accuracy but very likely requires more time.

C. More Details on Experiments

C.1. Datasets

The synthetic dataset is simulated with a physical particle-spring system, which contains five particles moving in a rectangular space. Some randomly selected pairs of particles are connected by invisible springs. The motion of particles are recorded as observations in the multivariate time series, including location and velocity. The motion of particles are determined by the laws of physics such as Newton's law, Hooke's law, and Markov property. The Connection dataset is perturbed by

re-sampling certain pairs of particles to be connected, therefore satisfy the requirement of correlation changes, as shown in (Zhang et al., 2020). The other two datasets (Location, Velocity) contain change points due to perturbation on the location of velocity therefore are different from Connection. In its original form, the synthetic datasets are used for supervised, offline setting, we adapt it into the unsupervised online setting in our experiments.

The Microservice dataset is generated from a Microservice platform consisting of six system nodes and 234 system pods. System faults on four dates are simulated system running for 49 hours. Specifically, six types of system metrics are recorded, (*i.e.*, CPU usage, memory usage, received bandwidth, transmit bandwidth, rate received packets, rate transmitted packets), and the time granularity of these system metrics is 1 second. This dataset has correlation-based changes because observations on system metrics are tempered by the underlying relationship between system nodes and pods, which may cause the system failure. The time point when the change of system status happens is recorded as a change point.

C.2. Experiment Details

The implementation of our algorithm is carried out with an Intel Core i9 CPU of 32GB memory, no GPU required. For the default setting in Table 2 the parameters are only adapted to different datasets but not optimized, the parameters of baseline methods are inherited from previous implementations. Recall that RIO-CPD has one "hard" parameter, the sliding window size and two "soft" parameters: threshold ρ in CUSUM and L indicating the lag between sampling data. For all datasets except HASC, we select $L = 1$, only because HASC has a long sequence and we make $L = 5$. The window size W is specified to be 10 for the Beedance dataset, 20 for Microservice and HASC, and 5 for all synthetic datasets. Finally, we present the average delay and computation time in Table 3.