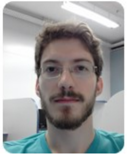
Non-parametric Online Change Point Detection on Riemannian Manifolds

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

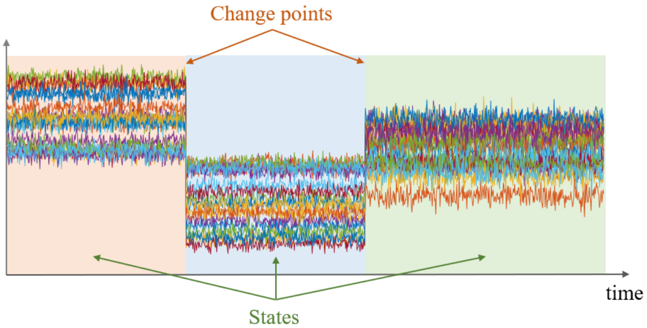
[†]Université Côte d'Azur, CNRS, OCA, France
[*]Université de Lorraine, CNRS, CRAN, France

To appear in ICML 2024.



Change point detection

Change point detection (CPD): detect abrupt changes in the states of time series[1],



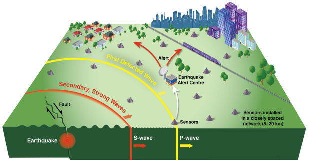
- non-parametric: no prior knowledge of the data distribution;
- online: process the stream on the fly, ideally without storing raw data.

[1] Samaneh Aminikhanghahi et al. "A survey of methods for time series change point detection". In: *Knowledge and information systems* 51.2 (2017), pp. 339–367.

CPD on Riemannian manifolds

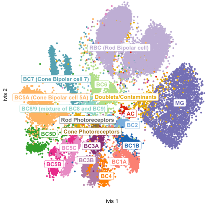
Riemannian manifold $\mathcal{M}$: curvature is induced by constraint, e.g., $\|\boldsymbol{x}\| = 1$ for the sphere, or metric, e.g., $\langle \xi_1, \xi_2 \rangle_{\boldsymbol{\Sigma}} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\xi_1\boldsymbol{\Sigma}^{-1}\xi_2)$ for the manifold of symmetric positive definite (SPD) matrices.

Many features of signals lie on manifolds, e.g., covariance descriptors and subspace representations. Investigate CPD on manifolds can impact many applications, e.g.,



earthquake detection[2]

video change detection[3]



subspace change detection[4]

[2] **https://www.earthquakescanada.nrcan.gc.ca/eew-asp/system-en.php**.

[3] **https://intvo.com/**.

[4] **https://bering-ivis.readthedocs.io/en/stable/**.

CPD on Riemannian manifolds

Developing methods on Riemannian manifolds is challenging:

- nonlinear geometry;
- lack of vector space structure.

Few works have investigated CPD for manifold-valued data:

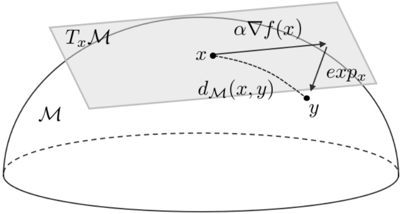
- parametric algorithm[5];
- offline technique[6].

This work introduces a general framework for non-parametric and online CPD on Riemannian manifolds.

[5] Florent Bouchard et al. "Riemannian geometry for compound Gaussian distributions: Application to recursive change detection". In: *Signal Processing* 176 (2020), p. 107716.

[6] Paromita Dubey et al. "Fréchet change-point detection". In: *The Annals of Statistics* 48.6 (2020), pp. 3312–3335.

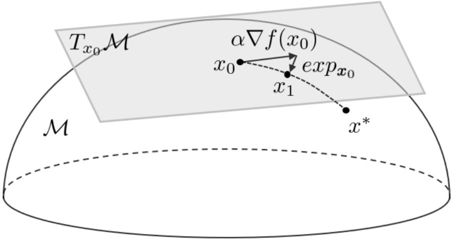
Riemannian optimization: main tools



A few important tools:

- Riemannian gradient: $\nabla f(x) \in T_x\mathcal{M}$;
- exponential mapping: $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ (maps a vector in the tangent space back to the manifold);
- Riemannian distance: $d_\mathcal{M}$ (length of the shortest path between two points on $\mathcal{M}$).

Riemannian optimization: R-SGD, basic structure



Considering a cost $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{M}$ we proceed as[7]:

- compute a stochastic approximation of $\nabla f(\boldsymbol{x})$ at $\boldsymbol{x}$;
- "take a step in the negative gradient direction" on $\mathcal{M}$ using the exponential mapping.

[7] Silvere Bonnabel. "Stochastic gradient descent on Riemannian manifolds". In: *IEEE Transactions on Automatic Control* 58.9 (2013), pp. 2217–2229.

Problem formulation

There exists a time index $t_r \in \mathbb{N}$ with an abrupt change in the probability measures[8] of $\boldsymbol{x}_t$ lying on $\mathcal{M}$, that is:

$$t < t_r : \ \boldsymbol{x}_t \sim P_1(\boldsymbol{x})\,, \qquad t \geq t_r : \ \boldsymbol{x}_t \sim P_2(\boldsymbol{x})\,, \tag{1}$$

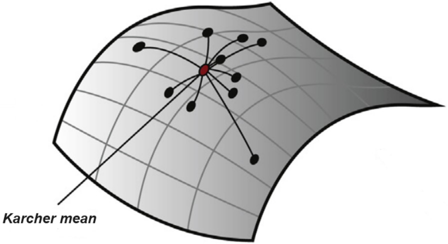
where t_r is the so-called *change point*.

The CPD problem on $\mathcal{M}$ consists of estimating t_r with the following requirements:

- high detection rate;
- low false alarm rate;
- low detection delay.

[8] Xavier Pennec. *Probabilities and statistics on riemannian manifolds: A geometric approach*. Tech. rep. 5093. INRIA, 2004, pp. 1–49.

The algorithm: the Karcher mean



Consider monitoring the Karcher mean[9] on $\mathcal{M}$, defined as

$$\boldsymbol{m}^* \in \arg\min_{\boldsymbol{m}} \ f(\boldsymbol{m}) \,. \tag{2}$$

where the Karcher variance

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

,

[9] Hermann Karcher. "Riemannian center of mass and mollifier smoothing". In: *Communications on pure and applied mathematics* 30.5 (1977), pp. 509–541.

The algorithm: online estimation

To achieve online detection, we consider using the R-SGD algorithm[10] to address problem (2):

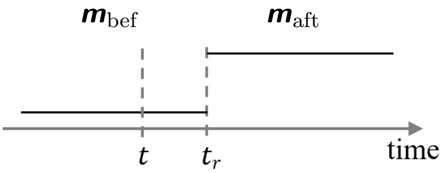
$$\boldsymbol{m}_{t+1} = \exp_{\boldsymbol{m}_t}\big(-\alpha H(\boldsymbol{m}_t, \boldsymbol{x}_t)\big), \tag{3}$$

where $H(\boldsymbol{m}, \boldsymbol{x})$ denotes the unbiased stochastic gradient of the loss such that

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

[10] Silvere Bonnabel. "Stochastic gradient descent on Riemannian manifolds". In: *IEEE Transactions on Automatic Control* 58.9 (2013), pp. 2217–2229.

The algorithm: an adaptive CPD



To detect change points by monitoring abrupt changes in $\boldsymbol{m}$,

- compute estimates $\widehat{\boldsymbol{m}}_{\text{bef}}$ and $\widehat{\boldsymbol{m}}_{\text{aft}}$;
- compare these two quantities using $d_{\mathcal{M}}(\widehat{\boldsymbol{m}}_{\text{bef}}, \widehat{\boldsymbol{m}}_{\text{aft}})$.

Rationale: the larger the $d_{\mathcal{M}}(\widehat{\boldsymbol{m}}_{\text{bef}}, \widehat{\boldsymbol{m}}_{\text{aft}})$, the more likely to flag t as a change point.

How to detect change points in an online way?

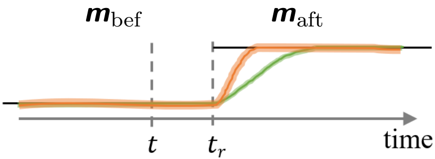
The algorithm: an adaptive CPD

We consider two estimates with two different fixed step sizes $\lambda < \Lambda$ as follows:

$$\boldsymbol{m}_{\lambda,t+1} = \exp_{\boldsymbol{m}_{\lambda,t}}\left(- \lambda H(\boldsymbol{m}_{\lambda,t}, \boldsymbol{x}_t)\right), \tag{4}$$

$$\boldsymbol{m}_{\Lambda,t+1} = \exp_{\boldsymbol{m}_{\Lambda,t}}\left(- \Lambda H(\boldsymbol{m}_{\Lambda,t}, \boldsymbol{x}_t)\right). \tag{5}$$

Convergence is directly affected by λ and Λ:



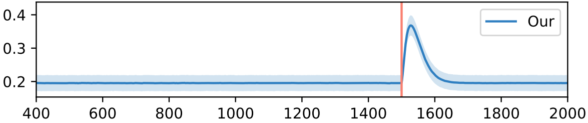
$\boldsymbol{m}_{\mathrm{bef}}$
$\boldsymbol{m}_{\mathrm{aft}}$
t
t_r
time

An adaptive CPD statistic is given by:

$$g_t = d_{\mathcal{M}}(\boldsymbol{m}_{\lambda,t}, \boldsymbol{m}_{\Lambda,t}). \tag{6}$$

CPD is then performed by comparing g_t to a threshold ξ.

The algorithm: preview of the results



1. Can we provide some performance guarantees?
2. How to determine a detection threshold ξ?

Theoretical analysis: convergence

The performance guarantee of our statistic g_t is based on a transient behavior in the convergence of the R-SGD algorithm:

Theorem

With some assumptions, for any $s \in \mathbb{N}_*$, the stochastic Riemannian gradient descent algorithm with a constant step size α satisfies:

$$\mathbb{E}\{f(\boldsymbol{m}_s) - f(\boldsymbol{m}^*)\} \leq \frac{(1-\epsilon)^{(s-1)} D^2}{2\alpha} + \frac{\alpha\sigma^2}{2\epsilon}\,, \tag{7}$$

with $\epsilon = \min\{\frac{1}{\zeta(\kappa,D)}, \alpha\mu\}$ and $\zeta(\kappa, D) = \frac{\sqrt{|\kappa|}D}{\tanh\left(\sqrt{|\kappa|}D\right)}$.

Theoretical analysis: performance guarantee

Theorem

Under the null hypothesis H_0, $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$ *are drawn i.i.d. from* $P(\boldsymbol{x})$ *with the Karcher mean* $\boldsymbol{m}^*$. *With some assumptions, at a steady state, the false alarm rate can be upper bounded by:*

$$\mathbb{P}\left(g_\infty \geq \xi \middle| \mathrm{H}_0\right) \leq \frac{2}{\xi}\left(f(\boldsymbol{m}^*) + \frac{(\lambda + \Lambda)\sigma^2}{4\epsilon}\right)^{\frac{1}{2}}, \tag{8}$$

with $\epsilon = \min\left\{\frac{1}{\zeta(\kappa, D)}, \lambda\mu\right\}$ *and* $\xi > 0$ *the detection threshold.*

A higher detection threshold ξ leads to a tighter bound, which is also influenced by Karcher variance $f(\boldsymbol{m}^*)$.

Theoretical analysis: performance guarantee

Theorem

Under the alternative hypothesis H_1, $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-B-1}$ *are drawn i.i.d. from* $P_1(\boldsymbol{x})$ *with Karcher mean* $\boldsymbol{m}_1^*$, *and* $\boldsymbol{x}_{t-B}, \boldsymbol{x}_{t-B+1}, \ldots, \boldsymbol{x}_{t-1}$ *are drawn i.i.d. from* $P_2(\boldsymbol{x})$ *with Karcher mean* $\boldsymbol{m}_2^*$. *With some assumptions, the detection rate can be lower bounded as:*

$$\mathbb{P}(g_t > \xi | \mathrm{H}_1) \geq \frac{d_{\mathcal{M}}(\boldsymbol{m}_1^*, \boldsymbol{m}_2^*) - \psi(\lambda) - \phi(\Lambda) - \xi}{D - \xi}, \tag{9}$$

where $\psi(\lambda) = \left(2 f_{\text{bef}}(\boldsymbol{m}_1^*) + \frac{\lambda\sigma^2}{\epsilon}\right)^{\frac{1}{2}} + \lambda\rho B$ *and* $\phi(\Lambda) = \left(2 f_{\text{aft}}(\boldsymbol{m}_2^*) + \frac{(1-\epsilon)^B D^2}{\Lambda} + \frac{\Lambda\sigma^2}{\epsilon}\right)^{\frac{1}{2}}$.

A lower detection threshold ξ leads to a tighter bound, which is also influenced by Karcher variances $f_{\text{bef}}(\boldsymbol{m}_1^*)$ and $f_{\text{aft}}(\boldsymbol{m}_2^*)$, and distance $d_{\mathcal{M}}(\boldsymbol{m}_1^*, \boldsymbol{m}_2^*)$.

Adaptive threshold selection

Under the null hypothesis, approximate g_t by a Gaussian distribution, set ξ as an estimate of the q-th quantile of g_t by computing only its first two moments[11]: $\beta_t^g = (1-\alpha)\beta_{t-1}^g + \alpha g_t$; $\gamma_t^g = (1-\alpha)\gamma_{t-1}^g + \alpha g_t^2$; $\hat{\xi}_t = \beta_t^g + \sqrt{\gamma_t^g - (\beta_t^g)^2}\sqrt{2}\mathrm{erf}^{-1}(2q-1)$.

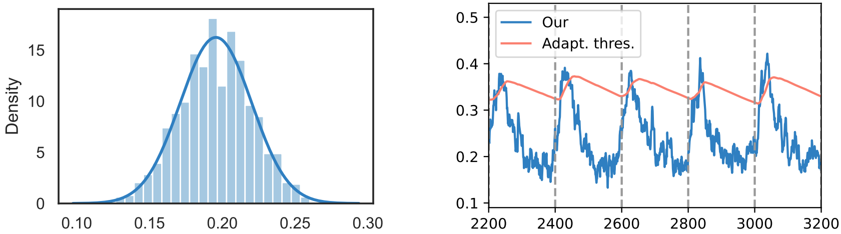


Figure: Distribution of g_t under the null hypothesis (left) and illustration of the adaptive threshold procedure (right).

[11] Nicolas Keriven et al. "NEWMA: a new method for scalable model-free online change-point detection". In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 3515–3528.

Applications and experiment setups

We apply our strategy to two manifolds as examples:

- the manifold of symmetric positive definite (SPD) matrices: $\mathcal{S}_p^{++}$;
- the Grassmann manifold: $\mathcal{G}_p^k$.

Baselines:

- Scan-B[12], NEWMA[13] and NODE[14]: designed for Euclidean spaces, online;
- F-CPD[15]: designed for manifold-valued data, offline.

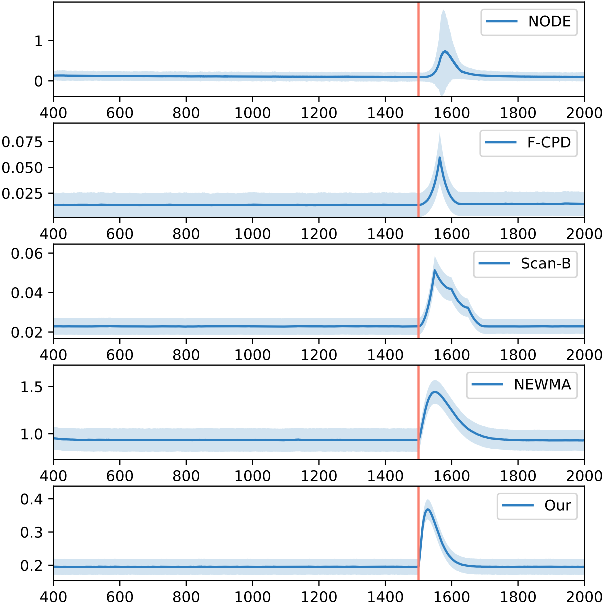
[12] Shuang Li et al. "Scan B-statistic for kernel change-point detection". In: *Sequential Analysis* 38.4 (2019), pp. 503–544.

[13] Nicolas Keriven et al. "NEWMA: a new method for scalable model-free online change-point detection". In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 3515–3528.

[14] Xiuheng Wang et al. "Change Point Detection with Neural Online Density-ratio Estimator". In: *IEEE international conference on acoustics, speech and signal processing (ICASSP)*. 2023.

[15] Paromita Dubey et al. "Fréchet change-point detection". In: *The Annals of Statistics* 48.6 (2020), pp. 3312–3335.

Experiment with synthetic data on $\mathcal{S}_p^{++}$



Experiment with synthetic data on $\mathcal{S}_p^{++}$

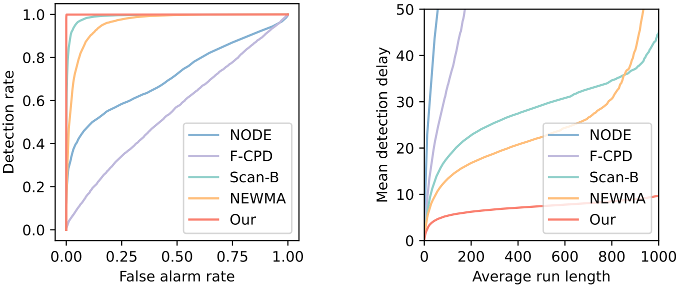
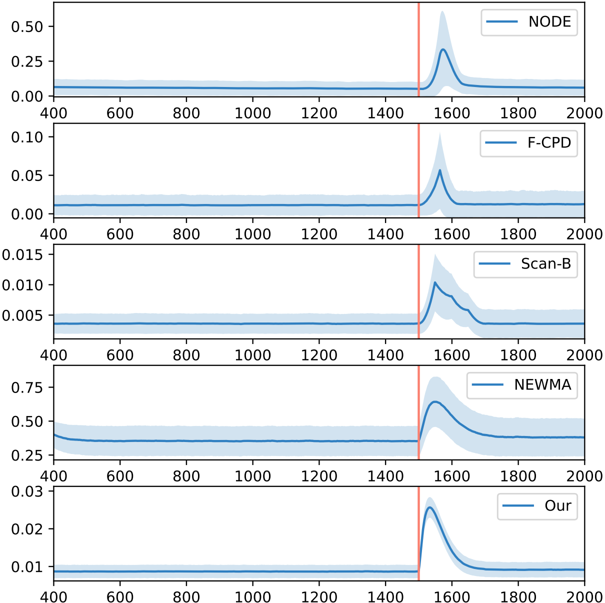


Figure: ROC curves, ARL versus MDD for the compared algorithms.

Experiment with synthetic data on $\mathcal{G}_p^k$



Experiment with synthetic data on $\mathcal{G}_p^k$

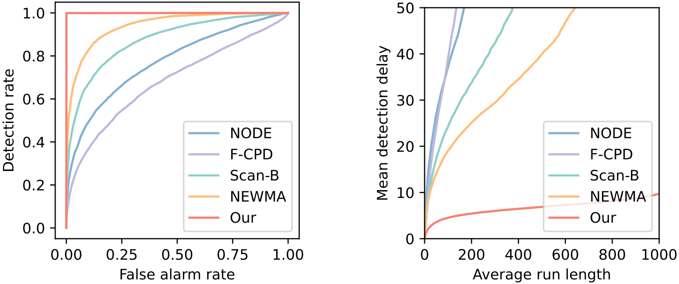


Figure: ROC curves, ARL versus MDD for the compared algorithms.

Voice activity detection

4 seconds of real speech from the TIMIT database[16] was added to 15 seconds of background noises from the QUT-NOISE database[17], with -3 dB Signal-to-Noise Ratio.

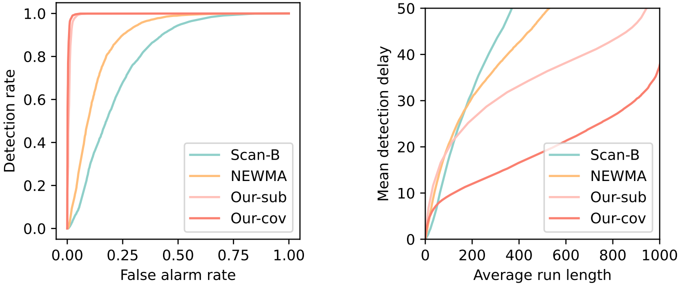


Figure: ROC curves, ARL versus MDD for voice action detection.

[16] John S Garofolo. “Timit acoustic phonetic continuous speech corpus”. In: *Linguistic Data Consortium, 1993* (1993).

[17] David Dean et al. “The QUT-NOISE-TIMIT corpus for evaluation of voice activity detection algorithms”. In: *Proceedings of the 11th Annual Conference of the International Speech Communication Association*. International Speech Communication Association. 2010, pp. 3110–3113.

Skeleton-based action recognition

Use the HDM05 motion capture database[18]. and generate data points $\mathbf{\Sigma}_t \in \mathcal{S}_p^{++}$ with $p = 93$ by computing the joint covariance descriptor[19] of 3D coordinates of the 31 joints.

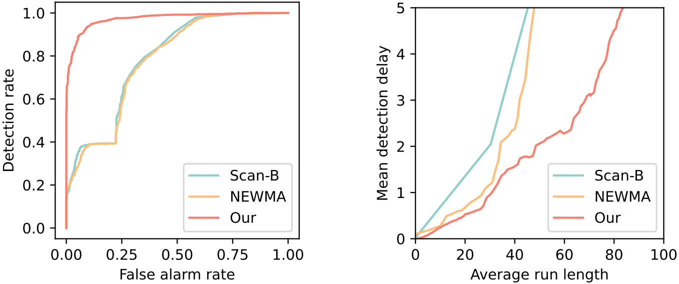


Figure: ROC curves, ARL versus MDD for skeleton-based action recognition.

[18] M. Müller et al. *Documentation Mocap Database HDM05*. Tech. rep. CG-2007-2. Universität Bonn, 2007.

[19] Mohamed E Hussein et al. "Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations". In: *Twenty-third international joint conference on artificial intelligence*. 2013.

Non-parametric Online Change Point Detection on Riemannian Manifolds

Xiuheng Wang, Ricardo Borsoi, Cédric Richard
xiuheng.wang@oca.eu
raborsoi@gmail.com
cedric.richard@unice.fr