Detecting change points in neural population activity with contrastive metric learning

Carolina Urzay[1,*], Nauman Ahad[1,*], Mehdi Azabou[1], Aidan Schneider[2], Geethika Atamkuri[1],
Keith B. Hengen[2], Eva L. Dyer[1]

1 - Georgia Institute of Technology,Atlanta, GA 30308 USA

2 - Washington University in St.Louis, St. Louis, MO 63130 USA

I. Abstract

Finding points in time where the distribution of neural responses changes (change points) is an important step in many neural data analysis pipelines. However, in complex and free behaviors, where we see different types of shifts occurring at different rates, it can be difficult to use existing methods for change point (CP) detection because they can't necessarily handle different types of changes that may occur in the underlying neural distribution. Additionally, response changes are often sparse in high dimensional neural recordings, which can make existing methods detect spurious changes. In this work, we introduce a new approach for finding changes in neural population states across diverse activities and arousal states occurring in free behavior. Our model follows a contrastive learning approach: we learn a metric for CP detection based on maximizing the Sinkhorn divergences of neuron firing rates across two sides of a labeled CP. We apply this method to a 12-hour neural recording of a freely behaving mouse to detect changes in sleep stages and behavior. We show that when we learn a metric, we can better detect change points and also yield insights into which neurons and sub-groups are important for detecting certain types of switches that occur in the brain.

> **Keywords:** Change Point Detection; Neural Population Activity; Metric Learning; Naturalistic Behavior Analysis; Contrastive Learning

II. Introduction

The brain is constantly in a state of flux, with variations occurring as individuals switch their attention to new tasks [12] or change their mood and overall state [16] Thus, in the analysis of data that spans multiple states or behaviors, identifying change points, or points in time where the distribution of neural responses shifts, becomes a critical task [8].

Change points have been studied in neural activity for single neurons [15] and in relatively simple tasks where there are a small number of pre-defined switches (e.g., Ready-Set-Go

has three phases and two change points) [10], [2]. However, there is a lot less work that examines the detection of changes in multi-neuron recordings during naturalistic behavior, where many different types of CPs can occur without any prompt or warning. In these cases, we need methods that can detect any number of diverse types of shifts that occur in the brain.

In this work, we establish a new metric learning approach for detecting change points in neural population activity. Our method uses a triplet loss to learn a metric where samples on different sides of a labeled change point are repelled from one another and nearby points on one side of a change point are brought closer (see Figure 1).

We use this approach to analyze a 12-hour block of neural activity from hippocampus (CA1) and show that we can better detect sleeping states and behavior from a small amount of labeled data. Critically, our method is also interpretable and can be used to reveal which neurons or interactions between neurons are important for revealing certain types of shifts in the population states. Our results suggest that by coupling metric learning with a contrastive sampling mechanism, we can build interpretable measures of shifts in brain state over long time periods.

Our contributions are as follows:

- We introduce a new and interpretable approach for change point detection in neural population activity.
- We show how multiple metrics can be learned jointly to facilitate change point detection in diverse settings.
- We apply our method to a 12 hour long recording from the mouse cortex and show that our approach can improve sleep stage recovery by identifying change points in neural activity more accurately.

III. Background

A. Change point detection

Change points are instances in a sequence where there is a change in the data generating distributions. In general online methods for change point detection (CPD), a measure of divergence (i.e., KL-divergence or Wasserstein distances) are often used to determine if two subsequent windows of data are generated from the same distribution or not [6], [4]. This measure of divergence is applied on two windows, a past window and a future window, which are slid along the length of the sequence. A change point is detected at instances where

the measure of divergence between future and past windows is greater than a set threshold.

Formally, a CP is detected at time t when the divergence d between the past window X_p^t and the future window X_f^t is greater than a set threshold τ. Thus, the decision rule can be expressed with the following equation:

$$d(X_p^t, X_f^t) \geq \tau.$$

Here, we will use this general online approach for CP detection in neural data. However, we will also show how using some supervised data to *learn a metric* can help to better detect change points of interest and provide added interpretability. This metric can be used and plugged into our Sinkhorn divergence and then used in an online manner.

B. Wasserstein distances and Sinkhorn divergences

Wasserstein distances compute the minimal cost of transporting mass from one distribution to another. Concretely speaking, consider two discrete multivariate distributions α and β on $\mathbb{R}^d$. We can express these distributions as

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \quad \text{and} \quad \beta = \sum_{j=1}^{m} b_j \delta_{y_j}, \qquad (1)$$

where δ_x is the Dirac function at position $x \in \mathbb{R}^d$, and similarly δ_y is the Dirac function at position $y \in \mathbb{R}^d$. Thus x_i and y_j denote the mass locations, while $a_i, b_i \in \mathbb{R}_+$ are the weights at these mass locations for the distributions α and β respectively.

The ground cost metric $C \in \mathbb{R}^{n \times m}$ represents the transportation cost between each pair of distribution mass locations. In this work, we consider **Wasserstein 2** ($\mathcal{W}^2$) distances that use a squared distance ground cost metric, where the $(i,j)^{\text{th}}$ entry of C is given by

$$C_{i,j} = \|x_i - y_j\|_2^2.$$

As the goal is to minimize the cost of moving mass between two distributions, Wasserstein distances require computing a transport plan P that dictates how mass is transported between the distributions. This is done by solving the following optimization problem:

$$\mathcal{W}(\alpha, \beta) = \min_P \langle C, P \rangle,$$
$$\text{subject to } P \in \mathbb{R}_+^{n \times m}, P^T \mathbb{1}_n = b, P \mathbb{1}_m = a,$$

where $\langle C, P \rangle$ is the Frobenius inner product between the cost matrix C and the transport plan P, a and b contain the mass weights for the distributions α and β, and $\mathbb{1}_n \in \mathbb{R}^n$ is the vector of all ones.

Wasserstein distances can be unstable and computationally expensive to compute, requiring $O(n^3 \log n)$ computations to evaluate in the case where n and m are of the same order. This makes it difficult to use Wasserstein distances repeatedly in two-sample tests. One solution to these problems is to add a regularization term $H(P)$ to form the entropic regularized Wasserstein distance $\mathcal{W}_\gamma^l$ [6], [14]. This is also known as the Sinkhorn distance and is defined as

$$\mathcal{W}_\gamma^l(\alpha, \beta) = \min_P \ \langle C, P \rangle - \gamma H(P), \qquad (2)$$
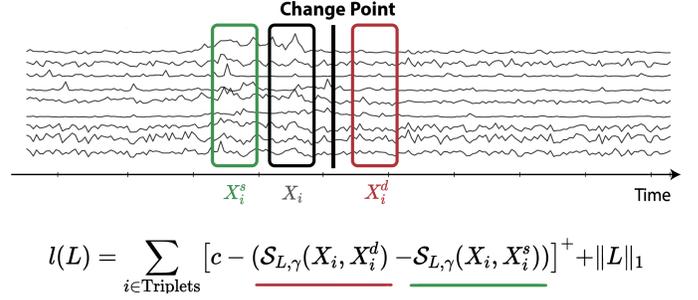$$\text{s.t } P \in \mathbb{R}_+^{n \times m}, P^T \mathbb{1}_n = b, P \mathbb{1}_m = a, \qquad (3)$$

$$l(L) = \sum_{i \in \text{Triplets}} \left[c - (\mathcal{S}_{L,\gamma}(X_i, X_i^d) - \mathcal{S}_{L,\gamma}(X_i, X_i^s)) \right]^+ + \|L\|_1$$

Fig. 1. *Obtaining triplet pairs from a true change point.* Sub-sequences X_i and X_i^s lie on the same side of the true change point shown by a vertical black line. These sub-sequences and are taken as a similar pair, while sub-sequences X_i and X_i^d lie on the opposite sides of this change point and are taken as a dissimilar pair. This provides a triplet pair (X_i, X_i^s, X_i^d)) which can be used to learn a ground metric parameter L

where $H(P)$ is the entropy of the transport plan matrix P and is given by

$$H(P) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(\log P_{i,j} - 1),$$

while γ is a regularization parameter. This regularization terms makes the minimization problem strongly convex and makes it less sensitive to changes in input, and can be solved with $O(n^2)$ computations using the Sinkhorn algorithm [6].

The regularized Wasserstein distance is biased as $\mathcal{W}_\gamma^2(\alpha, \alpha) \neq 0$. An unbiased divergence can be constructed from these regularized Wasserstein distances and is called the **Sinkhorn divergence**:

$$S_\gamma(\alpha, \beta) = \mathcal{W}_\gamma^l(\alpha, \beta) - \frac{1}{2}\mathcal{W}_\gamma^l(\alpha, \alpha) - \frac{1}{2}\mathcal{W}_\gamma^l(\beta, \beta). \qquad (4)$$

IV. METHODS

A. Learning a ground metric for optimal transport

While the squared distance is a natural choice for the ground cost metric C in (2) , available side information can also be used to learn a more suitable ground metric. This idea was first explored to directly estimate the ground cost using similarity/dissimilarity information for nearest neighbor classification tasks [7]. Similarity/dissimilarity information was also used to learn a Mahalanobis ground metric to compare word embeddings through Wasserstein distances in [9].

We leverage this idea of learning a ground metric for improving change point detection by using a recently developed method called SinkDivLM [1]. This method uses available change points to learn a ground metric C_L where the $(i,j)^{\text{th}}$ entry of this metric is given by,

$$C_{Li,j} = \|L(x_i - y_j)\|_2^2,$$

where L is a learnable parameter. This learned ground ground metric C_L can be used in place of C in (5) to obtain $S_{L,\gamma}$, which is Sinkhorn divergence equipped with this learned metric. SinkDivLM learns this parameter L by first obtain triplet pairs from available true change points. Sub-sequences on the same side of these change point are considered to be similar pairs while sub-sequences on opposite sides of change points are

considered to be dissimilar pairs. These similar, dissimilar pairs are used to obtain triplets (X_i, X_i^s, X_i^d). Figure 1 provides an example of how a true change point can be used to obtain a triple pair. These pairs are then used to learn a sparse metric L by minimizing

$$l(L) = \sum_{i \in \text{Triplets}} \left[c - (\mathcal{S}_{L,\gamma}(X_i, X_i^d) - \mathcal{S}_{L,\gamma}(X_i, X_i^s)) \right]^+ + \|L\|_1,$$
(5)

where $\mathcal{S}_{L,\gamma}$ is Sinkhorn divergence equipped with this learned metric, and c is triplet loss margin [1]. The L-1 regularization term helps learn a sparse metric which makes it easy to interpret what features or neurons are responsible for driving different types of changes.

B. Online approach for CP detection at test time

After learning the ground metric, we then can use an *online* or streaming approach for finding CPs. Concretely speaking, we can compute a change point statistic at time instance t by computing the Sinkhorn divergence $\mathcal{S}_{L,\gamma}$ between instance specific past and future windows X_p^t and X_f^t respectively. By streaming through the data and repeatedly computing this divergence on time instance specific past and future windows, we can generate change point statistics at all time instances within a stream. These change point statistics tell us the likelihood of detecting a change point. A change point is detected at instances where these statistics are greater than a set threshold τ. In addition, working on small windows locally at test time makes our approach inherently scalable to larger datasets.

C. Learning multiple metrics across diverse state changes

As our dataset consists of distinct types of sleep state changes, where each of these changes persists for different time scales and is triggered by different neurons, it is difficult to learn a single sparse metric that captures many diverse kinds of CPs while also providing interpretability. To solve this, we learn two metrics, one for sleep/wake transitions and one for REM/nREM transitions, and combine the two sparse metrics at inference time through simply pooling and re-normalizing both learned metrics. This highlights the flexibility of the approach in building in different types of labeled changes.

V. EXPERIMENTS

A. Dataset description

To study change point detection over long timescales and in naturalistic settings, we curated a dataset containing a 12 hour recording of behavior and neural activity data from the hippocampus of a mouse during free behavior [3]. We spike-sorted the neural data with MountainSort [5], then binned 42 neurons using 4-second windows resulting in a total of 10,800 samples of the population firing rates.

During the experiment, the animal moves in and out of different sleeping stages and natural behaviors. We performed sleep-scoring to obtain arousal states (wake, sleep REM and sleep nREM) using the recorded local field potentials [13]. We also annotated the discrete behaviors of the mouse through a manual tagging and analysis of the in-cage video (running, moving in place, standing still) that is recorded simultaneously.

B. Results

To show the improvements due to our metric learning approach, we compare our method with SinkDiv, a baseline method that uses the Sinkhorn divergence without a learned metric. Since change point detection performance is dependent on detection threshold, we use area under the curve (AUC) as an evaluation metric as it captures detection performance at different thresholds, and is commonly used as a change point evaluation metric in literature [11].

For this, we train the model on sleep/wake and REM/nREM labels, and we evaluate its performance with the corresponding labels, as well as all with the three of them combined (REM/nREM/wake).

The results in Table I show the mean AUC for the SinkDiv baseline and our model, SinkDivLM. These demonstrate that learning a metric is an improvement over the baseline, since the score increases from 0.58 to 0.85 in when trained and tested with sleep/wake labels and from 0.92 to 0.95 in the case of REM/nREM. As well, the combination of these metrics raise the score for identifying the changes among the 3 arousal substates, REM/nREM/wake.

Interestingly, SinkDivLM trained on sleep/wake labels does not outperform SinkDiv when tested with REM/nREM/wake labels but in the case of REM/nREM it does. A possible reason for this is that the neurons responsible for REM/nREM changes (6 and 15, as seen in Figure 2A) are shadowed by others in the sleep/wake metric (e.g. we can see neuron 6 is not highlighted in Figure 2B, while others are), which makes the detection of REM/nREM CPs a challenging task, as the values of these corresponding neurons are less high in comparison with the rest. In the opposite case, the learned metric for REM/nREM attributes high relative importance to neurons 13 and 15, which are also relevant for the detection of sleep/wake changes, being one of the brightest values in the matrix.

TABLE I
AUC OF CHANGE POINT DETECTION FOR SINKDIV BASELINE AND SINKDIVLM MODEL REPORTED ON TEST SETS WITH DIFFERENT TRUTH CP LABELS.

	SinkDiv	SinkDivLM
Trained on sleep/wake		
Sleep/wake	0.58	0.85
REM/nREM/wake	0.79	0.72
Trained on REM/nREM		
REM/nREM	0.92	0.95
REM/nREM/wake	0.79	0.82
Combined sleep metrics		
REM/nREM/wake	0.79	0.85
Trained on running/no running		
Running/no running	0.51	0.65

To check whether the learnt Sinkhorn divergence metric could identify changes not only in arousal states but also in complex behavior, a high level behavior label (running) was isolated. The model was trained on the 3 points where the mouse started or stopped running (running/no running). In
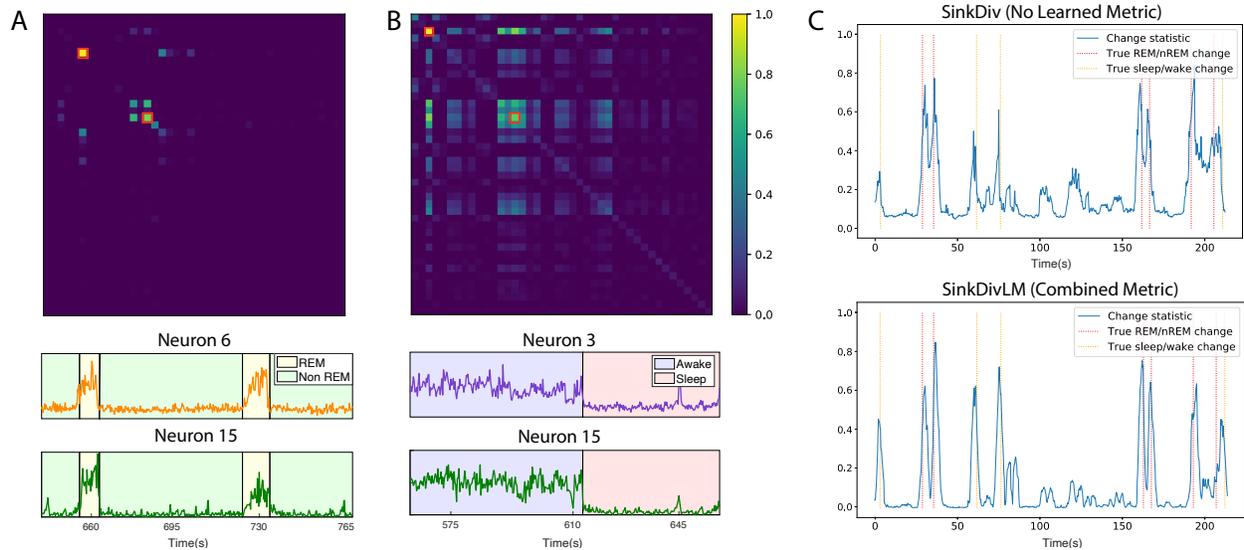
Fig. 2. *Summary of results*. We show the learned metrics for the sleep substates (REM/nREM on top A, sleep/wake on top B) and the firing rate of the most relevant neurons in each (highlighted in red in learned metric above) at examples of change points (bottom A, bottom B). C shows the change statistics for SinkDiv (top) SinkDivLM when metrics from both sleep/wake and REM/nREM are combined (bottom).

Table I, we visualize how, as in the case with sleep CPs, SinkDivLM reaches a much higher AUC than SinkDiv.

In addition to improving CP detection, our approach can be used to provide insight into neurons and subgroups of neurons that are most relevant for the detection of different CPs. In our analysis of sleep and wake states, we found that in the detection of nREM from REM sleep relies on a very small number of neurons overall. In contrast, detecting wake vs. sleep has more complex interactions across different neurons that contribute to the detection of these more macroscale arousal states. These results point to the utility of our approach, and its ability to lend itself to interpretability of changes in neural population states.

VI. CONCLUSION

In this paper, we developed a new approach for neural change point detection that uses a small amount of labeled changes to learn an interpretable metric to apply to neural populations. We show that contrastive metric learning can be used to improves the performance of change point detection in two kinds of neural shifts: arousal states and natural behaviors. We tested our approach on a large-scale neural recording spanning 12 hours with natural behavior and the animal moving in and out of sleep and wake freely and without prompt.

Such a tool has the potential to infer which behavioral changes are encoded in certain areas of the brain. These findings suggest a promising new direction towards building a tool for neuroscientists to analyze changes in neural activity during complex behavior.

REFERENCES

[1] Nauman Ahad, Eva L Dyer, Keith B Hengen, Yao Xie, and Mark A Davenport. Learning sinkhorn divergences for supervised change point detection. *arXiv preprint arXiv:2202.04000*, 2022.
[2] Bastian Alt, Michael Messer, Jochen Roeper, Gaby Schneider, and Heinz Koeppl. Non-parametric bayesian inference for change point detection in neural spike trains. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 258–262, 2018.
[3] Mehdi Azabou, Mohammad Gheshlaghi Azar, Ran Liu, Chi-Heng Lin, Erik C. Johnson, Kiran Bhaskaran-Nair, Max Dabagia, Bernardo Avila-Pires, Lindsey Kitchell, Keith B. Hengen, William Gray-Roncal, Michal Valko, and Eva L. Dyer. Mine your own view: Self-supervised learning through across-sample prediction, 2021. arxiv:2102.10106 [cs.LG].
[4] Kevin C Cheng, Shuchin Aeron, Michael C Hughes, Erika Hussey, and Eric L Miller. Optimal transport based change point detection and time series segment clustering. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Virtual Conference, 2020.
[5] Jason E. Chung, Jeremy F. Magland, Alex H. Barnett, Vanessa M. Tolosa, Angela C. Tooker, Kye Y. Lee, Kedar G. Shah, Sarah H. Felix, Loren M. Frank, and Leslie F. Greengard. A fully automated approach to spike sorting. *Neuron*, 95(6):1381–1394.e6, 2017.
[6] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Lake Thaoe, Nevada, 2013.
[7] Marco Cuturi and David Avis. Ground metric learning. *J. Mach. Learn. Res. (JMLR)*, 15(1):533–564, 2014.
[8] Arthur R Houweling and Michael Brecht. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451(7174):65–68, 2008.
[9] Gao Huang, Chuan Quo, Matt J Kusner, Yu Sun, Kilian Q Weinberger, and Fei Sha. Supervised word mover's distance. In *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Barcelona, Spain, 2016.
[10] Lena Koepcke, Go Ashida, and Jutta Kretzberg. Single and multiple change point detection in spike trains: Comparison of different cusum methods. *Frontiers in Systems Neuroscience*, 10, 2016.
[11] Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Netw.*, 43:72–83, 2013.
[12] Thomas Zhihao Luo and John HR Maunsell. Attention can be subdivided into neurobiological components corresponding to distinct behavioral effects. *Proceedings of the National Academy of Sciences*, 116(52):26187–26194, 2019.
[13] Zhengyu Ma, Gina G. Turrigiano, Ralf Wessel, and Keith B. Hengen. Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron*, 104(4):655–664.e4, 2019.
[14] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.*, 11(5-6):355–607, 2019.
[15] Rama Ratnam, Jozien B.M. Goense, and Mark E. Nelson. Change-point detection in neuronal spike train activity. *Neurocomputing*, 52-54:849–855, 2003. Computational Neuroscience: Trends in Research 2003.
[16] Benjamin John Shannon, Ronny A Dosenbach, Yi Su, Andrei G Vlassenko, Linda J Larson-Prior, Tracy S Nolan, Abraham Z Snyder, and Marcus E Raichle. Morning-evening variation in human brain metabolism and memory circuits. *Journal of neurophysiology*, 109(5):1444–1456, 2013.