Modeling Changes in Molecular Dynamics Time Series as Wasserstein Barycentric Interpolations

Jovan Damjanovic
Department of Chemistry
Tufts University

Yu-Shan Lin*
Department of Chemistry
Tufts University

James M. Murphy*
Department of Mathematics
Tufts University

***Abstract*—Molecular dynamics (MD) simulations are a powerful computational tool for elucidation of molecular behavior. These simulations generate an abundance of high-dimensional time series data and parsing these data into a human-interpretable format is nontrivial. Clustering trajectory *segments* obtained via change point detection has been shown to lower memory complexity and yield improved partitioning resolution of the time series compared to the state of the art. However, accurate change point placement is often inhibited by the presence of gradual changes between long-lived metastable states. The trajectory regions corresponding to these gradual changes are not well-modeled by a single distribution, and therefore are frequently over-segmented. In this work, we model such regions using weighted Wasserstein barycentric interpolations between adjacent metastable states, allowing for gradual changes to be resolved correctly. The improved detection performance of our proposed method is demonstrated on a range of toy and real MD simulation data, showing significant potential for faithfully modeling and compressing complex MD simulations.**

I. Introduction

Molecular dynamics (MD) simulations have long been a useful tool for scientists to explain the behavior of molecular systems that may not lend themselves well to experimental study [1]. Recent strides in both hardware and software have drastically increased the utility of MD as both an explanatory tool and a predictive one [1]–[3]. MD trajectories of even large biomolecular systems showcasing phenomena on millisecond timescales are now available, providing unique insights into fundamental problems such as protein folding [4]–[6], protein–ligand binding [7], and solution structural ensembles of macrocycles [1], [8], [9]. Such extensive trajectories, however, frequently consist of millions of data points across thousands of dimensions (corresponding to the positions in $\mathbb{R}^3$ of the constituent atoms or "intrinsic coordinates" thereof). To that end, tools such as dimensionality reduction [10], [11], cluster analysis [12]–[14], and Markov state modeling [15], [16] have become essential tools in computational chemists' arsenal. A brief overview of MD is provided in the Supplement.

Change point detection (CPD) has previously been shown to be an invaluable addition to this toolset [12], [17]. After using CPD to split MD trajectories into segments, subsequent cluster analysis performed using Wasserstein distances between segments benefits from improved clustering resolution, accounts for degrees of freedom omitted from the dataset, and yields fuzzy state boundaries which permit state overlap [12]. The preservation of temporal information by ensuring that data points that are close in time are given the same state assignments unless they are on opposite sides of a change point also allows for highly accurate transition modeling using Markov state modeling.

A key challenge is that, while MD simulations do yield *metastable states*—conformational clusters that correspond to relatively coherent fluctuations around a fixed shape—they also exhibit *transitions* between such states. Accounting for these transitions—which, like the metastable states themselves, are stochastic—is important for efficiently modeling MD simulations. In this sense, a reduced-order statistical model for MD simulations must capture not only metastable states, but gradual transitions between these states as well. An intuitive characterization of these gradual transitions is that of a "moving weighted average" between neighboring metastable states, wherein early points of a transition region are drawn from a distribution more similar to the initial state, and the later points are drawn from a distribution more similar to the final state.

To operationalize this characterization, we propose to model MD simulations as generated from metastable distributions and *Wasserstein barycentric interpolations* between them. This allows the metastable distributions to be interpolated in a geometry-preserving manner and our approach lends itself well to efficient implementations that can handle large MD simulations. In the remainder of the paper, we propose, justify, and analyze a specific choice of parametric model for the metastable states and resulting labeling algorithm (Section II) and validate our *Barycentric Transition (BarT)* approach on several synthetic and real MD simulation datasets (Section III) before concluding (Section IV).

II. Methods

Background and Motivation: Let $\mathcal{P}_{2,ac}(\Omega)$ denote the set of absolutely continuous probability measures on $\Omega \subset \mathbb{R}^D$ with finite second moment. Under the *instantaneous-transition* model, MD simulation trajectories $\{X(t)\}_{t=1}^T \subset \mathbb{R}^D$ are modeled as samples from a small collection of probability distributions $\{\nu_k\}_{k=1}^K \subset \mathcal{P}_{2,ac}(\mathbb{R}^D)$, each corresponding to a particular metastable state, intuitively a low-variance distribution around a canonical shape conformation. Equivalently, a metastable state can be characterized as a free energy basin

* Co-corresponding authors: yu-shan.lin@tufts.edu, jm.murphy@tufts.edu

whose depth governs the relative population of its corresponding state, whose width is related to the variance of the associated distribution, and with the height of the free energy barriers to the remaining states accounting for the lifetime of the state. The existence of a small number of metastable states captures the intuition that the molecule has a small number of canonical shapes, and data observed from the simulation are small fluctuations around these shapes, with relatively infrequent and instantaneous transitions between shapes.

Per this model, a simulation trajectory may be split into segments such that each segment is drawn from a single distribution corresponding to one of the metastable states. It is possible to model these distributions nonparametrically, but in practice it is common to set each ν_k to be a *Laplace distribution*. The wide tails of a Laplace distribution are representative of the noise patterns observed in MD simulations, [17] and the simple parametric form allows for relatively efficient estimation. In modelling MD trajectories as *mixtures of Laplace distributions*, the switch from one distribution to another occurs at change points $\tau_1 < \cdots < \tau_M$ which are unknown but which can be learned.

However, the instantaneous-transition model breaks down in most real-life MD simulation trajectories, wherein transitions often occur on timescales slower than the sampling rate necessary to observe chemically relevant phenomena. Since these transition regions are not sampled from a single metastable distribution (e.g., a fixed Laplace distribution), we propose to model MD simulation trajectories as consisting of metastable states *and transitions between them*. In our model, metastable distributions are still modeled as distinct Laplace distributions. The transition regions are modeled as samples from the measures constituting the *geodesic between the bounding metastable distributions in Wasserstein-2 space* [18]. More precisely, for two distributions $\mu_0, \mu_1 \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$, the *Wasserstein-2 geodesic* between them is the path $\gamma : [0,1] \to \mathcal{P}_{2,ac}(\mathbb{R}^d)$ such that $\gamma(t) = (tT_{\mu_0 \to \mu_1} + (1-t)\,\mathrm{Id})_{\#}\mu_0$, where $T_{\mu_0 \to \mu_1} = \underset{T_{\#}\mu_0 = \mu_1}{\arg\min} \int_{\mathbb{R}^D} \|x - T(x)\|_2^2 d\mu_0$ is the optimal transport (Monge) map between μ_0 and μ_1 [19]. We note that Wasserstein geodesics are special cases of *Wasserstein barycenters* and Laplace distributions are preserved under barycentric combinations [20]; see the Supplement for details.

In this context, we model our data $\{X(t)\}_{t=1}^T$ as being either (i) a sample from a *metastable* distribution (modeled as a Laplace distribution) or (ii) a *transition* region (a barycentric interpolation of previous and subsequent metastable distributions). Change points $1 < \tau_1 < \cdots < \tau_M < T$ denote switches from a metatable state to a transition region or another metastable state. So, for time $t \in [\tau_j, \tau_{j+1}]$, $X(t)$ is either sampled from a Laplace distribution with fixed parameters or from the Wasserstein barycenter between previous and subsequent Laplace distributions ν_{j-1} and ν_{j+1} with weights $\left(\frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}, 1 - \frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}\right)$ for $t \in \{\tau_j, \tau_j + 1, \ldots, \tau_{j+1}\}$.

Algorithm Overview: Change detection for this model is the key computational problem, which we proceed to address by recognizing that Wasserstein barycenters between Laplace distributions in a single dimension have a simple closed form. We learn change points via a penalized likelihood estimation approach, which aims to maximize

$$\mathcal{L}(\{\tau_j\}_{j=1}^M) = \sum_{j=1}^{M} L(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}) - \lambda \sum_{j=1}^{M} |S_j|^{\alpha}. \quad (1)$$

Here, L is the log-likelihood of the trajectory segments being described by the proposed model, including all segments in the trajectory, but with each dimension in a multivariate time series being considered separately; S_j is the number of change points detected across the D variables at time τ_j; and $\alpha \in (0,1], \lambda > 0$ are tunable parameters. We define $L(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}) = \max\{L_{\text{ms}}(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}), L_{\text{tr}}(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}})\}$ where L_{ms} and L_{tr} are log-likelihood functions corresponding to metastable and transition regions, respectively. We model the metastable states as Laplace distributions and the transition regions as Wasserstein barycenters between previous and subsequent Laplace distributions, leading to explicit log-likelihood functions as follows. If μ_j and σ_j are the mean and standard deviation of segment $\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}$, respectively, then $L_{\text{ms}}(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}) = \log \prod_{t=\tau_j}^{\tau_{j+1}} \frac{1}{2\sigma_j} \exp\left(-\frac{|X(t)-\mu_j|}{\sigma_j}\right)$, $L_{\text{tr}}(\{X(t)\}_{t=\tau_j}^{\tau_{j+1}}) = \log \prod_{t=\tau_j}^{\tau_{j+1}} \frac{1}{2\sigma_{j,t}} \exp\left(-\frac{|X(t)-\mu_{j,t}|}{\sigma_{j,t}}\right)$, where $\mu_{j,t} = \frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}\mu_{j-1} + \frac{t-\tau_j}{\tau_{j+1}-\tau_j}\mu_{j+1}, \sigma_{j,t} = \frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}\sigma_{j-1} + \frac{t-\tau_j}{\tau_{j+1}-\tau_j}\sigma_{j+1}$ denote the mean and standard deviation of the Laplace distribution that is the barycenter between the $\text{Lap}(\mu_{j-1}, \sigma_{j-1})$ and $\text{Lap}(\mu_{j+1}, \sigma_{j+1})$ with weights $\left(\frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}, 1 - \frac{\tau_{j+1}-t}{\tau_{j+1}-\tau_j}\right)$; see Propositions 1 and 2 in the Supplement for explicit derivations.

The second term in (1) is a penalty applied to each change point to prevent overfitting. The tuning parameter $\lambda > 0$ determines the sensitivity of the algorithm: a higher λ value will result in fewer changes being detected [17]. Furthermore, applying chemical intuition which suggests that important conformational changes are typically characterized by concurrent changes in multiple variables [21], the penalty may be adjusted so that changes occurring in multiple variables at once are less severely penalized: a lower value of α will result in a lower penalty for simultaneous change points [17]. See the Supplement for details on optimizing (1) and pseudocode below for implementation details.

III. Experiments and Analysis

Data Set Description: To gauge performance with respect to gradual transitions, we consider three data sets[1].

(i) A family of *toy-models* which are two-state, 1-D, noisy time series consisting of 20 "metastable" segments, alternating between two Laplace distributions. The segment lengths were generated randomly, with a minimum segment length of 100. The standard deviation of the two metastable distributions was kept constant within a time series, and was otherwise varied between 0 and 20, in increments of 1. The metastable

[1] All of the data sets and code are publicly available [22].

Algorithm 1 BarT pseudocode.

for $t = 1 \rightarrow T - 1$ **do**
  for $j = 1 \rightarrow J$ **do**
    $rand[j,t] \leftarrow Unif(0.9, 1)$
    $penalty[j,t] \leftarrow$
$(\lambda q(\{1, ..., J\}) - \lambda q(\{1, ..., J\} \backslash \{J\})) \cdot rand[j,t]$
  end for
end for
$shift\_and\_merge \leftarrow$ False
loop
  $changes \leftarrow \{\}$
  for $j = 1 \rightarrow J$ **do**
    $changes\_1D[j] \leftarrow$
optimize_1D($\{Y_{j,1}, ..., Y_{j,T}\}, penalty[j,:]$)
    for $t \in changes\_1D[j]$ **do**
      append($changes[t], j$)
    end for
  end for
  if $shift\_and\_merge$ **then**
    $(\tau_1, ..., \tau_k) \leftarrow$ sorted keys of $changes$
    for $i = 1 \rightarrow K$ **do**
      $L \leftarrow -\inf$
      for $s = \tau_{i-1} \rightarrow \tau_{i+1}$ **do**
        $l \leftarrow$ updated penalized LL, greater of the
"metastable" and "transition" LLs
        if $l > L$ **then**
          $L \leftarrow l$ and label the segment as
metastable or transition based on which l was kept
          $\tau_i^* \leftarrow s$
        end if
      end for
      $\tau_i \leftarrow \tau_i^*$
      shift or merge $changes[\tau_i]$ to $changes[\tau_i^*]$
    end for
  end if
  if $changes == changes\_prev$ **then**
    return $changes$
  end if
  if count($changes$) == $count\_prev$ **then**
    $shift\_and\_merge \leftarrow$ True
  end if
  for $t = 1 \rightarrow T - 1$ **do**
    for $j = 1 \rightarrow J$ **do**
      if $t \in changes$ **then**
        if $j \in changes[t]$ **then**
          $penalty[j,t] \leftarrow$
$\lambda q(changes[t]) - \lambda q(changes[t] \backslash \{j\})$
        else
          $penalty[j,t] \leftarrow$
$\lambda q(changes[t] \bigcup \{j\}) - \lambda q(changes[t])$
        end if
      else
        $penalty[j,t] \leftarrow \lambda q(\{j\}) \cdot rand[j,t]/0.9$
      end if
    end for
  end for
end loop

segments were separated by transition segments, with length of these segments kept constant within a time series, and varying between 1 and 100, in 1-point increments, across the different trajectories. These transition segments are generated from the Laplace barycenters corresponding to the weights and starting and ending Laplace distributions.

(ii) An MD-like synthetic trajectory with well-separated states and changes readily identifiable by visual inspection. We devised a *Langevin dynamics* example with a double-well quadratic potential, with wells centered at –1 and 1, respectively, and the potential barrier height set to 8. The potential curve is shown in the Supplement. Two 100,000 point trajectories were simulated using the molecular dynamics simulator OpenMM [23] with particle mass set to 1 amu, temperature set to 500 K, and damping coefficients of 10 ps^{-1} and 100 ps^{-1}, respectively.

(iii) *Alanine dipeptide*, a simple molecule consisting of a single amino acid, capped on both ends. It is a frequently used model system in theoretical chemistry [24]–[28]. We performed a 200 ns conventional MD simulation of alanine dipeptide using the GROMACS software suite [29] and the RSFF2 force field [30] in explicit TIP3P water [31]. The trajectory was sampled at a 1 ps sampling rate, yielding 200,000 data points. The trajectory was represented in terms of two *internal coordinates*—-the backbone dihedral angles ϕ and ψ—which have been shown to describe the essential motions of this molecule well [24], [25], [32] and can in fact be learned [10]. Of particular interest were the areas outside the canonical Ramachandran regions (see Supplement for a brief overview of Ramachandran plots) sampled in the trajectory, previously shown to be representative of gradual transitions [33], [34].

Results: The proposed algorithm, BarT (*Bar*ycentric *T*ransitions), was compared to the original implementation of SIMPLE [17] with parameters manually tuned to values yielding optimal results. The SIMPLE algorithm applies the same penalized likelihood estimation methodology as BarT; however, SIMPLE does not allow for transition regions to be captured in the likelihood estimation. Moreover, we compared BarT to a parametric global method—PELT [35]. This approach minimizes the total residual error with respect to a chosen test statistic. In order to give PELT the best chance of capturing transition regions, which we have already identified as "sloped", we defined the total deviation as the sum of squared differences between the data points and the predictions of the least-squares linear fit through the values. While PELT can also be configured using a sensitivity parameter—in this case, a threshold specifying the minimum decrease in residual error in order for a change point to be placed—it may also be configured with a user-specified number of change points to be detected. In order to give the competing algorithm an advantage, we allowed the number of change points to be equal to the ground truth, when applicable. With respect to the toy model systems with known ground truth, the algorithms were compared based on the number of change points detected (unless user-set) and their

placement accuracy. The Langevin dynamics dataset was compared by visual inspection, with change point placements overlaid on top of the time series view, and regions flagged as either metastable or transition, depending on which segment characterization maximizes the log-likelihood. On the alanine dipeptide dataset, the algorithms were compared both visually as described above, and based on the percentage of points belonging to noncanonical regions (as determined by cluster analysis [12]) flagged as belonging to transition regions.

On trajectories with exaggerated transition segments, barycentric modeling improves change detection accuracy and parameter robustness: The results of applying BarT, SIMPLE, and PELT to a representative synthetic trajectory are shown in Figure 1, with additional trajectory and tuning parameter combinations in the Supplement. Examining the results shows that with equivalent parameter tuning, SIMPLE is both more prone to over-segmentation and underestimates the width of the transition intervals. Moreover, upon calculating the log-likelihoods of the data being described by either a single Laplace distribution or a weighted barycentric interpolation between the Laplace distributions corresponding to adjacent trajectory segments, and labeling the segments accordingly based on which of the two log-likelihoods is greater, it is readily apparent that BarT greatly outperforms SIMPLE in terms of labeling. Results on additional trajectories, shown in the Supplement, suggest that SIMPLE may only yield somewhat comparable labels in very well-behaved, low-noise time series, or time series with short transition lengths, and that even in those cases, BarT displays improved change point placement. This is due to the fact that SIMPLE continues to underestimate the width of transition segments as it attempts to "split the difference" and distribute parts of regions that do not conform to its single-distribution assumption between adjacent trajectory segments. PELT, on the other hand, performs comparably to BarT, with fewer misclassifications than SIMPLE. Note, however, that neither PELT nor SIMPLE *classify* the data points themselves; instead, the BarT classification is used in all cases. This classification may fail on very short (1–2 data points) transition segments, wherein the log-likelihood of a transition segment being described by the barycentric transition model does not exceed that of it being described by a single Laplace distribution even with correct change point placement. Artifacts arising from this issue may be seen in Figure 2, although on the whole, the results from all 2,100 trajectories, shown in this figure, confirm that BarT outperforms the other two algorithms with respect to classification. Overall, BarT is capable of capturing gradual changes in time series that fit the proposed model Laplace transition model.

BarT yields intuitive change point placement on synthetic Langevin trajectories: Simulating the behavior of a particle in a two-well potential landscape provides a convenient, one-dimensional, two-state model of an MD trajectory, which can be tuned in a way which enables CPD algorithm comparison by visual inspection. The comparative performance of the three tested algorithms on a 100,000-point trajectory is shown in Figure 3. When tuned for low sensitivity, all algorithms

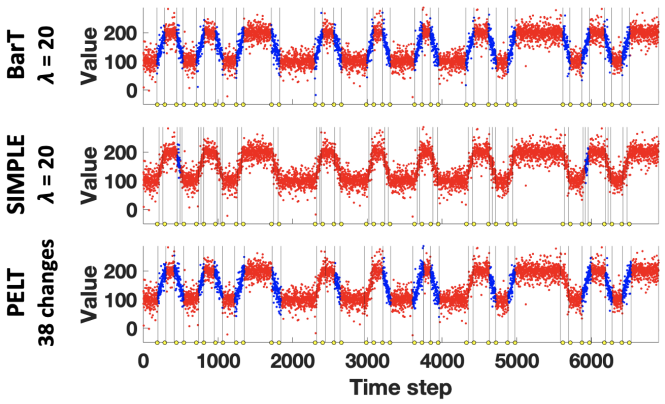

Fig. 1. Comparative performance of BarT, SIMPLE, and PELT on a noisy synthetic trajectory (μ_1 = 100, μ_2 = 200, σ = 20). Detected change points are marked with vertical lines. Segments labeled as metastable are shown in red, and segments labeled as transitions are shown in blue. Ground-truth change points are shown in yellow.

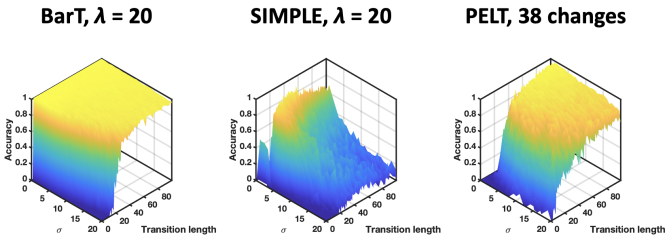

Fig. 2. Comparative performance of BarT, SIMPLE, and PELT on the entire set of noisy synthetic trajectories. Accuracy is defined as the fraction of transition points correctly identified.

prioritize high-amplitude changes, and BarT and PELT yield change point placements that result in the majority of transition regions being labeled correctly, whereas SIMPLE does not. The presence of noise in a trajectory contributes to imperfect labeling accuracy; this challenge may be addressed by using "soft" labeling instead, wherein the user may set the log-likelihood difference threshold to be accepted as a transition. When the algorithms are tuned for high sensitivity instead (shown in Supplement), we can observe that all methods treat both high-amplitude changes between the two states in the trajectory and local fluctuations within a state as change points. Again, only BarT and PELT are able to assign change points in a manner which results in gradual transition between states being correctly labeled as such. It is apparent that the MD-like trajectory contains transition features that do not always perfectly fit the model (e.g., incomplete or reversed transitions, or transitions via a longer-lived intermediate state). Regardless, BarT treats the vast majority of transition regions correctly, suggesting it may provide a way to reliably indicate the presence of gradual transitions in a trajectory.

BarT correctly localizes transition segments in alanine dipeptide, but noise prevents high sensitivity: The results of applying the three algorithms to the alanine dipeptide MD trajectory are shown in Figure 4. In the case of SIMPLE (Figure 4A, mid-left), once the relative log-likelihood labeling

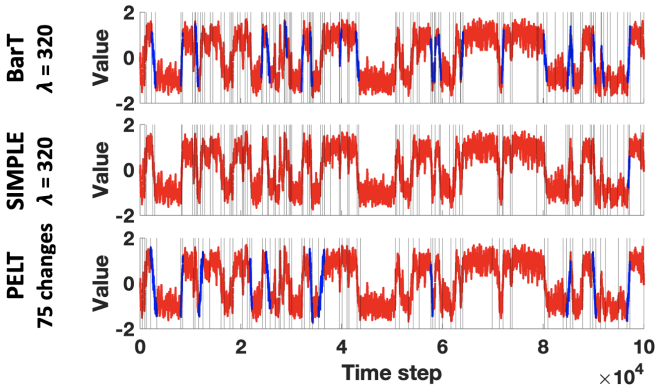

Fig. 3. Comparative performance of BarT, SIMPLE, and PELT on a Langevin dynamics trajectory under a two-well potential, in a low sensitivity regime (λ = 320 for SIMPLE and BART, 75 changes for PELT). Detected change points are marked with vertical lines. Segments labeled as metastable are shown in red, and segments labeled as transitions are shown in blue.

scheme is applied, 439 transition points are identified in ϕ, and 1,313 transition points are identified in ψ. Intuitively, we would expect transitions in ϕ to appear along horizontal paths connecting the two high-density regions with high ψ values: β and PPII (see Supplement for details). On the other hand, transitions in ψ may be expected to appear along vertical paths connecting the β and PPII regions to the α region underneath. Whereas this expectation is generally met, many putative transition points are identified in areas which, by visual inspection, do not correspond to transition regions, and which, by chemical intuition, are inside the canonical Ramachandran regions. In addition, a high number of points outside of the canonical regions remain unlabeled as transitions. Meanwhile, PELT labels 4,450 points in the ϕ time series as transitions, along with 4,760 points in the ψ time series, achieving a high rate of detection. Among these points, points outside the canonical regions are over-represented, in line with chemical intuition. BarT produces significantly different results (Figure 4A, mid-right): the very noisy ϕ time series (shown in Supplement) is labeled as having 278 transition points, which appear to correspond to local fluctuations, rather than large-scale gradual transitions. In the ψ time series, 6,175 transition points are identified. These points appear to fall into two main categories: those that do, in fact, correspond to "sloped" regions in the time series, and which are predominantly located outside the canonical Ramachandran regions, along vertical paths connecting β and PPII regions to the α region, and those belonging to long, flat segments split in a way which allows the central portion to be interpreted as a barycentric interpolation between the outside portions (Figure 4B). This error leads to a relatively high number of false positives, which may be addressed by tuning the λ parameter value further to avoid excessive splitting. Similarly to what we have observed in the case of SIMPLE, a high number of points outside the canonical areas remain unlabeled. This observation corroborates our findings on the toy datasets, which suggest that noise may lower the sensitivity of the algorithm. "Soft" labeling may be useful in this case as well. Note, while alanine dipeptide is an extremely common model system, due to its small size, it exhibits very rapid conformational transitions – larger systems with slower dynamics may benefit from the application of BarT more. Lastly, the application of BarT's classification in conjunction with algorithms like PELT, whose test statistic appears more robust to noise in this context, may be a worthwhile pursuit.

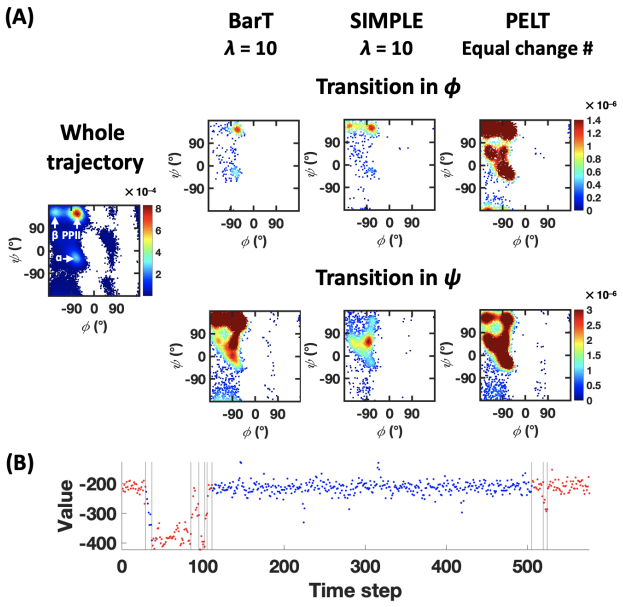

Fig. 4. Comparative performance of BarT, SIMPLE, and PELT on a 200 ns conventional MD trajectory of alanine dipeptide. (A) Density plots of the full trajectory and transition points in each variable, detected by each algorithm. (B) BarT false positive: central segment of a split "flat" region is treated as a barycentric interpolation between the outside segments.

IV. Conclusion

In this work, we presented a CPD algorithm which applied the penalized likelihood estimation principle to a model consisting of Laplace-distributed segments and barycentric interpolations thereof. Results on model systems show that this way of treating gradual transitions is effective and allows for more information to be obtained from MD and MD-like trajectories than using the current state-of-the-art algorithms which assume instantaneous transitions.

BarT is parametric, because it imposes Laplace distributions on the metastable states. It is of interest to extend the proposed framework and related ones based on Gaussian mixtures [36], [37] to the nonparametric setting for richer analysis of MD simulation data. Moreover, the proposed approach only allows for barycenters between two probability distributions. It is of interest to leverage recent work characterizing barycenters between more than two probability measures [38] to allow for efficient Wasserstein dictionary learning [39]–[41] in the context of MD simulations.

Acknowledgements: The authors acknowledge the Dreyfus Program for Machine Learning in the Chemical Sciences Award and a Tufts University DISC Seed Grant. We thank Matt Werenski for helpful conversations.

REFERENCES

[1] J. Damjanovic, J. Miao, H. Huang, and Y.-S. Lin, "Elucidating solution structures of cyclic peptides using molecular dynamics simulations," *Chemical Reviews*, vol. 121, no. 4, pp. 2292–2324, 2021.

[2] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 120–127, 2009.

[3] J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa, M. Feig, and Y. Sugita, "New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems," *Journal of Computational Chemistry*, vol. 42, no. 4, pp. 231–241, 2021.

[4] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, "Millisecond-scale molecular dynamics simulations on anton," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ser. SC '09. New York, NY, USA: Association for Computing Machinery, 2009.

[5] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins," *Science*, vol. 330, no. 6002, pp. 341–346, 2010.

[6] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *Science*, vol. 334, no. 6055, pp. 517–520, 2011.

[7] A. Li and M. K. Gilson, "Protein-ligand binding enthalpies from near-millisecond simulations: Analysis of a preorganization paradox," *The Journal of Chemical Physics*, vol. 149, no. 7, p. 072311, 2018.

[8] H. Geng, F. Chen, J. Ye, and F. Jiang, "Applications of molecular dynamics simulation in structure prediction of peptides and proteins," *Computational and Structural Biotechnology Journal*, vol. 17, pp. 1162–1170, 2019.

[9] A. M. Razavi, W. M. Wuest, and V. A. Voelz, "Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models," *Journal of Chemical Information and Modeling*, vol. 54, no. 5, pp. 1425–1432, 2014.

[10] M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, "Determination of reaction coordinates via locally scaled diffusion map," *The Journal of chemical physics*, vol. 134, no. 12, p. 03B624, 2011.

[11] W. Zheng, M. A. Rohrdanz, M. Maggioni, and C. Clementi, "Polymer reversal rate calculated via locally scaled diffusion map," *The Journal of chemical physics*, vol. 134, no. 14, p. 144109, 2011.

[12] J. Damjanovic, J. M. Murphy, and Y.-S. Lin, "Catboss: Cluster analysis of trajectories based on segment splitting," *Journal of Chemical Information and Modeling*, vol. 61, no. 10, pp. 5066–5081, 2021.

[13] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[14] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham, "Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms," *Journal of Chemical Theory and Computation*, vol. 3, no. 6, pp. 2312–2334, 2007.

[15] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Current Opinion in Structural Biology*, vol. 25, pp. 135–144, 2014.

[16] B. E. Husic and V. S. Pande, "Markov state models: From an art to a science," *Journal of the American Chemical Society*, vol. 140, no. 7, pp. 2386–2396, 2018.

[17] Z. Fan, R. O. Dror, T. J. Mildorf, S. Piana, and D. E. Shaw, "Identifying localized changes in large systems: Change-point detection for biomolecular simulations," *Proceedings of the National Academy of Sciences*, vol. 112, no. 24, pp. 7454–7459, 2015.

[18] R. McCann, "A convexity principle for interacting gases," *Advances in mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[19] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[20] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM Journal on Mathematical Analysis*, vol. 43, no. 2, pp. 904–924, 2011.

[21] H. Huang, J. Damjanovic, J. Miao, and Y.-S. Lin, "Cyclic peptides: backbone rigidification and capability of mimicking motifs at protein–protein interfaces," *Phys. Chem. Chem. Phys.*, vol. 23, pp. 607–616, 2021.

[22] https://github.com/ysl-lab/bart-change-detection.

[23] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLOS Computational Biology*, vol. 13, pp. 1–17, 2017.

[24] V. Spiwok and P. Kříž, "Time-lagged t-distributed stochastic neighbor embedding (t-SNE) of molecular simulation trajectories," *Frontiers in Molecular Biosciences*, vol. 7, 2020.

[25] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, "Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach," *Chemical Physics Letters*, vol. 509, no. 1, pp. 1–11, 2011.

[26] W. Zheng, M. A. Rohrdanz, and C. Clementi, "Rapid exploration of configuration space with diffusion-map-directed molecular dynamics," *The Journal of Physical Chemistry B*, vol. 117, no. 42, pp. 12 769–12 776, 2013.

[27] W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. J. C. Ward, Y. Zhestkov, and R. Zhou, "Describing protein folding kinetics by molecular dynamics simulations. 2. example applications to alanine dipeptide and a β-hairpin peptide," *The Journal of Physical Chemistry B*, vol. 108, no. 21, pp. 6582–6594, 2004.

[28] Y. Zhao, F. K. Sheong, J. Sun, P. Sander, and X. Huang, "A fast parallel clustering algorithm for molecular simulation trajectories," *Journal of Computational Chemistry*, vol. 34, no. 2, pp. 95–104, 2013.

[29] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19–25, 2015.

[30] C.-Y. Zhou, F. Jiang, and Y.-D. Wu, "Residue-specific force field based on protein coil library. RSFF2: Modification of AMBER ff99sb," *The Journal of Physical Chemistry B*, vol. 119, no. 3, pp. 1035–1047, 2015.

[31] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[32] A. Ma and A. R. Dinner, "Automatic method for identifying reaction coordinates in complex systems," *The Journal of Physical Chemistry B*, vol. 109, no. 14, pp. 6769–6779, 2005.

[33] S. Hayward, "Peptide-plane flipping in proteins," *Protein Science*, vol. 10, no. 11, pp. 2219–2227, 2001.

[34] R. Caliandro, G. Rossetti, and P. Carloni, "Local fluctuations and conformational transitions in proteins," *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4775–4785, 2012.

[35] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[36] K. Cheng, S. Aeron, M. C. Hughes, and E. L. Miller, "Dynamical Wasserstein barycenters for time-series modeling," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 991–28 003, 2021.

[37] K. C. Cheng, S. Aeron, M. C. Hughes, and E. L. Miller, "Non-parametric and regularized dynamical Wasserstein barycenters for time-series analysis," *arXiv preprint arXiv:2210.01918*, 2022.

[38] M. Werenski, R. Jiang, A. Tasissa, S. Aeron, and J. M. Murphy, "Measure estimation in the barycentric coding model," in *International Conference on Machine Learning*. PMLR, 2022, pp. 23 781–23 803.

[39] J. Bigot, R. Gouet, T. Klein, and A. López, "Geodesic PCA in the Wasserstein space by convex PCA," in *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, vol. 53, no. 1, 2017, pp. 1–26.

[40] M. A. Schmitz, M. Heitz, N. Bonneel, F. Ngole, D. Coeurjolly, M. Cuturi, G. Peyré, and J.-L. Starck, "Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 643–678, 2018.

[41] M. Mueller, S. Aeron, J. M. Murphy, and A. Tasissa, "Geometric sparse coding in Wasserstein space," *arXiv preprint arXiv:2210.12135*, 2022.

[42] M. Knott and C. S. Smith, "On the optimal mapping of distributions," *Journal of Optimization Theory and Applications*, vol. 43, pp. 39–49, 1984.

[43] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[44] R. Stafford, "Random Vectors with Fixed Sum", (https://www.mathworks.com/matlabcentral/fileexchange/9700-random-

vectors-with-fixed-sum), MATLAB Central File Exchange, 2006. Retrieved October 5, 2021.

[45] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *Journal of Molecular Biology*, vol. 7, no. 1, pp. 95–99, 1963.