Learning neuronal manifolds for interacting neuronal populations

Akshey Kumar (akshey.kumar@univie.ac.at)
Research Group Neuroinformatics, Faculty of Computer Science, University of Vienna, Austria

Moritz Grosse-Wentrup (moritz.grosse-wentrup@univie.ac.at)
Research Group Neuroinformatics, Faculty of Computer Science, University of Vienna, Austria

Abstract

Understanding how neuronal populations interact to process information and generate behavior is a central goal of neuroscience. However, high dimensionality, dense interactions, and unobserved factors complicate this task. The neuronal manifold hypothesis suggests that relevant dynamics occur on a lower-dimensional manifold, but it offers limited insight into the interactions among subsystems. We introduce a BunDLe-Net-based architecture that embeds distinct neuronal populations into separate latent dimensions. By leveraging BunDLe-Net's Markovian embedding, we ensure that every point in the latent space retains predictive information about future behavioral dynamics. We apply our method to *C. elegans* neuronal data categorized into sensory, motor, and interneurons. The manifold not only reveals recurring motifs in the dynamics but also shows how different populations orchestrate these motifs. From the manifold, we can read off which populations encode information and drive the dynamics in each behavioral state. Thus, we present a powerful visual tool that reveals how information is processed and relayed across populations.

Keywords: neuronal manifolds; population dynamics; behavior; modular neural systems; neural representations.

Introduction

Advances in data collection techniques have significantly improved our ability to record brain activity at unprecedented scales. However, transforming this raw data into insights about how the brain processes information is challenging and requires extensive analysis. Recently, neuronal manifold learning has emerged as a powerful approach for interpreting high-dimensional neuronal recordings (Mitchell-Heggs, Prado, Gava, Go, & Schultz, 2023). These methods map neuronal data onto low-dimensional manifolds for improved interpretability.

While some approaches utilize standard dimensionality reduction techniques (Cunningham & Yu, 2014; Gao et al., 2017), others specifically address the time-series nature of neuronal data, embedding it in a way that respects temporal dynamics (Kumar, Gilra, Gonzalez-Soto, Meunier, & Grosse-Wentrup, 2024; Schneider, Lee, & Mathis, 2023; Pandarinath et al., 2018). Unlike most methods that reconstruct or simulate neuronal activity, BunDLe-Net selectively discards information irrelevant to behavioral dynamics (Kumar et al., 2024). This yields a minimal behavioral model that reveals bundled trajectories, where the branching structure of dynamics reflects behavioral motifs. This branching is unique to BunDLe-Net, as it isolates relevant dynamics while filtering out noise.

Despite these advances, current methods fail to capture the interplay of neuronal populations and how they interact to process information. Most existing methods embed all neurons jointly in a latent space, leading to entangled representations that obscure the distinct roles of individual populations. Those that do learn disentangled representations, do so without modeling interactions between the populations (Kobak et al., 2016; Miller, Eckstein, Botvinick, & Kurth-Nelson, 2024). To address these limitations, we propose an approach that endows each latent dimension with interpretability by associating it with a known subsystem of interest. Specifically, we embed the neurons from each subsystem into distinct dimensions of the latent space, while allowing for temporal interactions between them. This separation allows us to disentangle the contributions of each subsystem, providing a clearer understanding of how different populations interact to process information and orchestrate behavior.

We apply our method to calcium imaging data from *C. elegans* neurons, recorded alongside behavioral data (discrete behavior based on the locomotor state of the worm) (Kato et al., 2015). For our analysis we utilise neuron categories from (Kaplan, Nichols, & Zimmer, 2018) and partition the identified neurons into three subsets: $\mathcal{X}^{(1)}$ for sensory neurons, $\mathcal{X}^{(2)}$ for interneurons, and $\mathcal{X}^{(3)}$ for motor neurons.

Architecture for learning Markovian representations

Let X_t represent the global neuronal state at time t, and define a mapping $\tau : X_t \mapsto Y_t$, where Y_t is a lower-dimensional, coarser latent representation of X_t. Let T_Y denote the temporal transition model at in the latent space. BunDLe-Net learns a

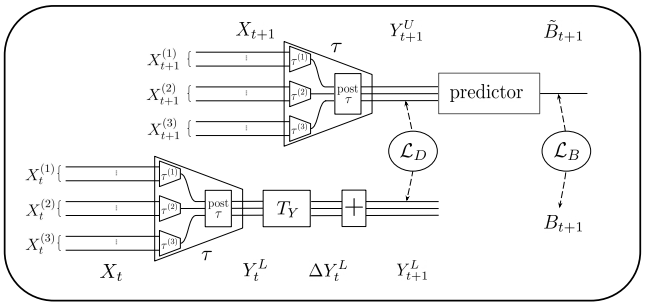

Figure 1: BunDLe-Net architecture for interacting populations

latent representation that 1) preserves behavioral information 2) preserves dynamical information pertaining to the behavior, and in doing so respects the temporal causality of the system (Grosse-Wentrup, Kumar, Meunier, & Zimmer, 2023). In the neural network architecture, 1) is achieved by requiring behavior B to be decodable from Y through the *predictor* layer. 2) is achieved by requiring that embedding the time-evolved state $X_{t+1} \xrightarrow{\tau} Y_{t+1}$ or time evolving the embedding $X_t \xrightarrow{\tau} Y_t \xrightarrow{T_Y} Y_{t+1}$ yield the same result. This is enforced by upper and lower arm of architecture in Figure 1 respectively.

The original BunDLe-Net architecture performed the embedding τ jointly on all neurons. In this work, to study interacting populations, we embed each of subsystems separately to a distinct dimension of latent space. We modularize the τ layer to incorporate three non-overlapping mappings denoted by $\tau^{(1)}$, $\tau^{(2)}$ and $\tau^{(3)}$. These mappings operate on the vectors $X^{(1)}$, $X^{(2)}$, and $X^{(3)}$ respectively. The same three $\tau^{(i)}$ are ap-

plied to both X_t and X_{t+1}, ensuring a consistent embedding across time steps. Note that, though we embed subsystems separately, we later pass the embedding to a joint transition model. This allows our embedding to model interactions between the subsystems.

Results

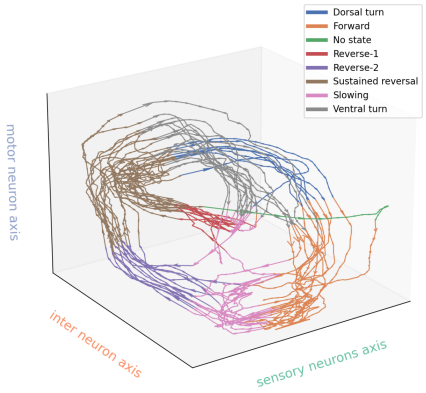

Figure 2: BunDLe-Net embedding of *C. elegans* neuronal data where each dimension corresponds to a specific subset of neurons. The behavioral state is denoted by color.

Figure 2 shows the embedding of worm neuronal activity in a three-dimensional[1] latent space where each of the axes correspond to a neuronal population. The manifold shows a branching topology, indicating that the BunDLe-Net architecture has successfully abstracted the neuronal representation of behavior. Note that while the behavior is discrete, the neurons encode it in a continuous representation that is unknown to the external behavior annotator. Since we embed each population to a separate dimension, the fact that the manifold is still three dimensional (rather than on a plane, line or point) indicates that all three populations globally encode for behavior and play a part in information processing. This aligns with existing knowledge in neuroscience concerning *C. elegans* neuronal circuitry (Kaplan et al., 2018).

We see that the trajectories are bundled together and, while they merge and branch, the bundles never intersect each another for different behaviors. This is on account of the Markovian embedding which ensures that maximum predictive information about both the behavior and its future state are preserved in this embedding.

From the manifold (Figure 2), we can read-off not only *which populations encode for behavior* but also *which populations drive the dynamics*. Firstly, note how certain behaviors are projected along a specific direction or confined to a given plane (see animated 3D plot). For example, we see that the slowing ● and sustained reverse ● are separated along interneurons axis. Secondly, notice how the trajectory changes its orientation with axes through time. For example, the reverse-1 ● and reverse-2 ● behaviors are primarily aligned along the interneuron axis, while the dorsal turn ● is initiated on the sensory neuron axis. To quantify how much different populations contribute to driving behavior, we present Figure 3

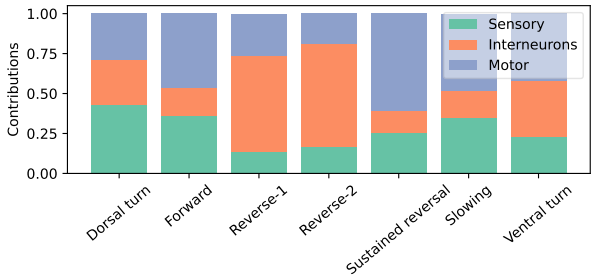

Figure 3: Populations that drive behavioral dynamics. From the BunDLe-Net trajectory, we see that certain behaviors show dynamics along specific directions. This indicates that certain populations are more responsible in driving a given behavior. We quantify this by looking at the absolute of the sum of direction vectors for a given behavior.

Discussion

The resulting manifolds allows us to make statements such as, *The dynamics is initiated in population $X^{(1)}$, which then relays the information to population $X^{(2)}$; the information is then shared with $X^{(3)}$ which jointly orchestrates the behavior B with $X^{(2)}$, and so on....* From such embeddings, one can thus read-off interactions between various subsystems, their relationships with one another, and the flow of information between them.

Our method can be of great practical significance to the experimentalist since the manifold can not only help deciding *which* populations to stimulate in order to induce a behavioral motif, but also *when* to stimulate. For example consider the branching point in the manifold at sustained reverse ●. Even before the future behaviors (dorsal ● and ventral turn ●) are externally observed, the trajectory bifurcates. This means the internal neuronal representation has already been set on a deterministic path to one of the behaviors. Thus one should ideally stimulate *at* the bifurcation and not after. We also see how dorsal turn initiation ● → ● plays out on the sensory axis, whereas the ventral turn initiation ● → ● initiation involves motor neurons.

[1] The dimensionality of three is specific to the BunDLe-Net method and always works for discrete behaviors. It is not related with the task dimensionality or the dimension of the null-space

Acknowledgments

This work has been funded by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/LS23070]

References

Cunningham, J. P., & Yu, B. M. (2014, August). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. Retrieved from http://dx.doi.org/10.1038/nn.3776 doi: 10.1038/nn.3776

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., & Ganguli, S. (2017). A theory of multineuronal dimensionality, dynamics and measurement. *bioRxiv*. doi: 10.1101/214262

Grosse-Wentrup, M., Kumar, A., Meunier, A., & Zimmer, M. (2023). Neuro-cognitive multilevel causal modeling: A framework that bridges the explanatory gap between neuronal activity and cognition. *bioRxiv*. doi: 10.1101/2023.10.27.564404

Kaplan, H. S., Nichols, A. L., & Zimmer, M. (2018, September). Sensorimotor integration in caenorhabditis elegans: a reappraisal towards dynamic and distributed computations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1758), 20170371. Retrieved from https://doi.org/10.1098/rstb.2017.0371 doi: 10.1098/rstb.2017.0371

Kato, S., Kaplan, H. S., Schrödel, T., Skora, S., Lindsay, T. H., Yemini, E., ... Zimmer, M. (2015, October). Global brain dynamics embed the motor command sequence of caenorhabditis elegans. *Cell*, *163*(3), 656–669.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., ... Machens, C. K. (2016, apr). Demixed principal component analysis of neural population data. *eLife*, *5*, e10989. Retrieved from https://doi.org/10.7554/eLife.10989 doi: 10.7554/eLife.10989

Kumar, A., Gilra, A., Gonzalez-Soto, M., Meunier, A., & Grosse-Wentrup, M. (2024). Bundle-net: Neuronal manifold learning meets behaviour. *bioRxiv*. Retrieved from https://www.biorxiv.org/content/early/2024/04/15/2023.08.08.551978 doi: 10.1101/2023.08.08.551978

Miller, K., Eckstein, M., Botvinick, M., & Kurth-Nelson, Z. (2024). Cognitive model discovery via disentangled rnns. *Advances in Neural Information Processing Systems*, *36*.

Mitchell-Heggs, R., Prado, S., Gava, G. P., Go, M. A., & Schultz, S. R. (2023). Neural manifold analysis of brain circuit dynamics in health and disease. *Journal of Computational Neuroscience*, *51*(1), 1–21.

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., ... Sussillo, D. (2018, September). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*(10), 805–815. Retrieved from https://doi.org/10.1038/s41592-018-0109-9 doi: 10.1038/s41592-018-0109-9

Schneider, S., Lee, J. H., & Mathis, M. W. (2023, May). Learnable latent embeddings for joint behavioural and neural analysis. , *617*(7960), 360–368. doi: 10.1038/s41586-023-06031-6