Across-animal odor decoding by probabilistic manifold alignment — Supplementary information—

Pedro Herrero-Vidal
Center for Neural Science
Neuroscience Institute
New York University
pmh314@nyu.edu

Dmitry Rinberg
Neuroscience Institute
Center for Neural Science
NYU Langone Health
rinberg@nyu.edu

Cristina Savin
Center for Neural Science
Center for Data Science
New York University
cs5360@nyu.edu

S1 amLDS inference: Kalman smoothing

The goal is to estimate the posterior distribution of the latent state $\mathbf{z}_t$ using all available measurements $\mathbf{x}_{1:T}$. This posterior has joint multivariate Gaussian statistics with marginal parameters $\mu_{t|T}$ and $\Sigma_{t|T}$, which is computed in two stages, a forward pass corresponding to Kalman filtering and a backward pass which implements Kalman smoothing. We will use the double index $t|t'$ to denote both the time index of the latent variable, t, and to signal the range of observations that the posterior is conditioned on t', namely, $\mathrm{P}(\mathbf{z}_t|\mathbf{x}_{1:t'}, \theta) = \mathcal{N}\left(\mu_{t|t'}, \Sigma_{t|t'}\right)$.

The Kalman filtering involves recursive updates in a forward pass, alternating between updates that propagate beliefs about the current state $\mathbf{z_t}$ given only the previous state $\mathbf{z_{t-1}}$, using the stimulus-specific parameters:

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{A}_k\boldsymbol{\mu}_{t|t-1} + \mathbf{b}_{k,t} \tag{1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{A}_k\boldsymbol{\Sigma}_{t|t-1}\mathbf{A}_k^\top + \mathbf{Q}_k, \tag{2}$$

and steps that incorporate evidence from the current observation $\mathbf{x_i}$, via the animal-specific observation model parameters:

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{x}_t - \mathbf{C}_m\boldsymbol{\mu}_{t|t-1}) \tag{3}$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} + \mathbf{K}_t\mathbf{C}_m\boldsymbol{\Sigma}_{t|t-1} \tag{4}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_m^\top(\mathbf{C}_m\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_m^\top + \mathbf{R})_m^{-1}, \tag{5}$$

where $\mathbf{K}_t$ denotes the Kalman gain.

The smoothing (backward pass) incorporates evidence from the later part of the observation sequence:

$$\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t} + \mathbf{F}_t(\mu_{t+1|T} - \boldsymbol{\mu}_{t+1|t}) \tag{6}$$

$$\boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|t} + \mathbf{F}_t(\Sigma_{t+1|T} - \boldsymbol{\Sigma}_{t_1|t})F_t^\top \tag{7}$$

$$\mathbf{F}_t = \boldsymbol{\Sigma}_{t|t}\mathbf{A}_k^\top\boldsymbol{\Sigma}_{t+1|t}^{-1}, \tag{8}$$

with $\mathbf{F}_t$ denoting the smoothing analogue of the Kalman gain.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).
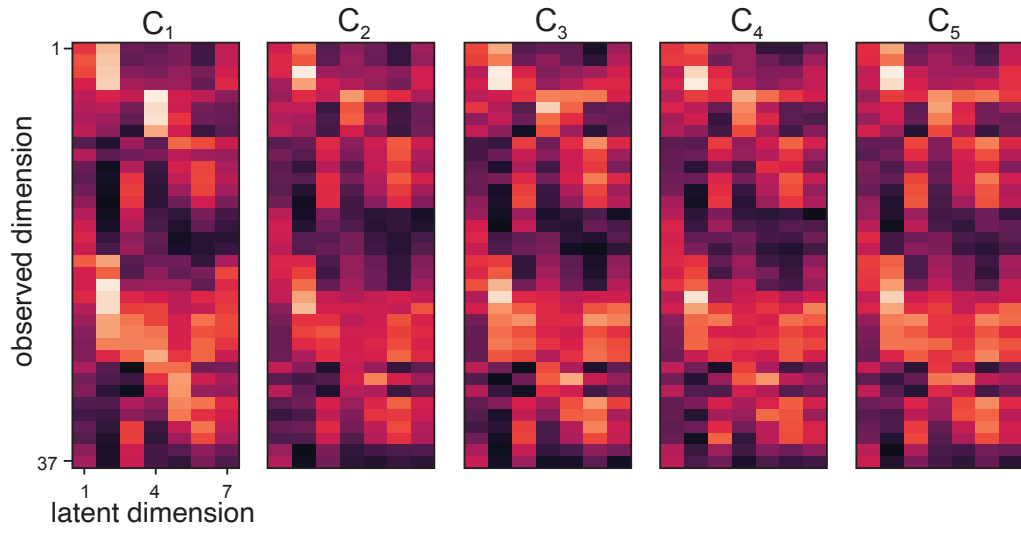
Figure S1: **Manifold alignment across animals.** Animal-specific loading matrices $\mathbf{C}_m$ estimated using amLDS from the olfactory bulb neural datasets. The observation space (y-axis) includes the set of electrodes that are active (not damaged) in all animals.

2

Checklist

1. For all authors...

   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Yes, it's accurate.

   (b) Did you describe the limitations of your work? [Yes] See Section 2 and Discussion.

   (c) Did you discuss any potential negative societal impacts of your work? [N/A] We don't foresee any obvious negative societal impact from this work.

   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] Yes, it conforms.

2. If you are including theoretical results...

   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 2.

   (b) Did you include complete proofs of all theoretical results? [Yes] See Section 2 and Supplementary 1.

3. If you ran experiments...

   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] A full description of the model and simulations is provided on Sections 2-4. Supplementary code is provided to replicate the results. A complete description of the neural data is provided on Section 4.

   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Sections 2-4

   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See figure captions and Sections 3 and 4

   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 3. amLDS memory and computational demands are really low, both parameter learning and inference can run on a laptop computer.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

   (a) If your work uses existing assets, did you cite the creators? [Yes] We use code from scikit-learn, which we credit in reference [27].

   (b) Did you mention the license of the assets? [N/A]

   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] The model and a user tutorial is provided as supplementary material.

   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

3