Interpreting dynamics of neural activity after dimensionality reduction

Timothée Proix[1], Matthew G. Perich[2], Tomislav Milekovic[3,4]

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland
[2] Department of Neuroscience, Université de Montréal, Montréal, Canada
[3] Defitech Center for Interventional Neurotherapies (NeuroRestore), Lausanne, Switzerland
[4] Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland

ABSTRACT

Recorded activity of cortical neurons during behavior is often analyzed with dimensionality reduction methods. The resulting oscillatory trajectories have been interpreted as signatures of latent oscillatory dynamical systems. Here we show that oscillatory trajectories arise as a consequence of the horseshoe effect after applying dimensionality reduction methods on signals that approximately exhibit continuous variation in time, regardless of whether latent oscillatory dynamical systems are present or not. We show that task-relevant information is instead contained in deviations between oscillatory trajectories, which we account for by introducing a new model that goes beyond the dichotomy between representational and dynamical systems models of neural activity. Our findings help interpret the results of dimensionality reduction analyses.

Main text

Modern experimental techniques allow researchers to monitor the activity of hundreds to thousands of neurons simultaneously. Analyzing, interpreting and understanding such high-dimensional neural activity is challenging. Dimensionality reduction methods are commonly used to try to help simplify the interpretation of this activity[1,2]. Consistent, interpretable low-dimensional structure in the identified trajectories is often taken as a meaningful insight into the underlying computations or mechanistic structure of the recorded neural population[3,4]. For example, dimensionality reduction methods have shown correlated patterns of activity that trace out oscillatory trajectories when applied to the neuronal activity of motor cortical neural populations during movement[3,5,6], to the neuronal activity of the prefrontal cortex during working memory[7], or to the LFP activity of the lateral sensorimotor cortex during speech articulation[8]. The shapes of these neural modes show stability across recorded neuron sets, tasks, trials and time[9]. These oscillatory trajectories have been interpreted as evidence that the activity of motor cortical populations behaves as a latent oscillatory dynamical system[3,4]. However, fields such as genetics[10], ecology[11], climatology[12], time series analysis[13] and neuroscience in some specific cases[14,15] have recognized the emergence of the "horseshoe effect" after applying numerous dimensionality reduction methods, including principal component analysis (PCA)[16] and multidimensional scaling[17]. This horseshoe effect corresponds to the oscillatory trajectories that appear when applying dimensionality reduction to data exhibiting continuous spatial or temporal variations. In this article, we first demonstrate that the horseshoe effect can apply to a large variety of neural population recordings. We introduce analytical derivations and simulations that provide a cautionary

tale on the risks of misinterpreting observed oscillatory structure. We show that oscillatory trajectories are frequently found in neural recordings, even when they largely depart from a homogeneous continuous variation in time, thus reflecting a generic feature of neural activations. We finally propose that informative information is instead contained in deviations of trajectories across conditions, and introduce a new model to bring new insights into the dimensionality and structure of the encoding space.

To illustrate how the horseshoe artifact emerges, consider a poster session at the 2022 Society for Neuroscience (SfN) meeting in San Diego. Thousands of neuroscientists throughout the world have gathered for a poster. We ask these neuroscientists to record the time on their watches across a half day (0 to 12 hours) as they attend each poster. Being here for the first time, most of them are excited and have forgot to adjust their watches to local Pacific Standard Time (PST), causing the measurements to differ by a phase shift given by the time difference between their home time zone and PST (Fig. 1a). We collect the measurements from the same number of neuroscientists from 12 time zones equally space in time and apply PCA on this dataset. Regardless of the spatial ordering of data, the PCA finds the orthogonal linear combinations of watch records (principal components, here called modes) that explain the variance in decreasing order. The first mode explains the most variance of the whole dataset, the second mode the most variance of the full dataset minus the first mode, etc. Intriguingly, we would find that the time evolution of the modes, called "latent variables", show consistent oscillatory structure.

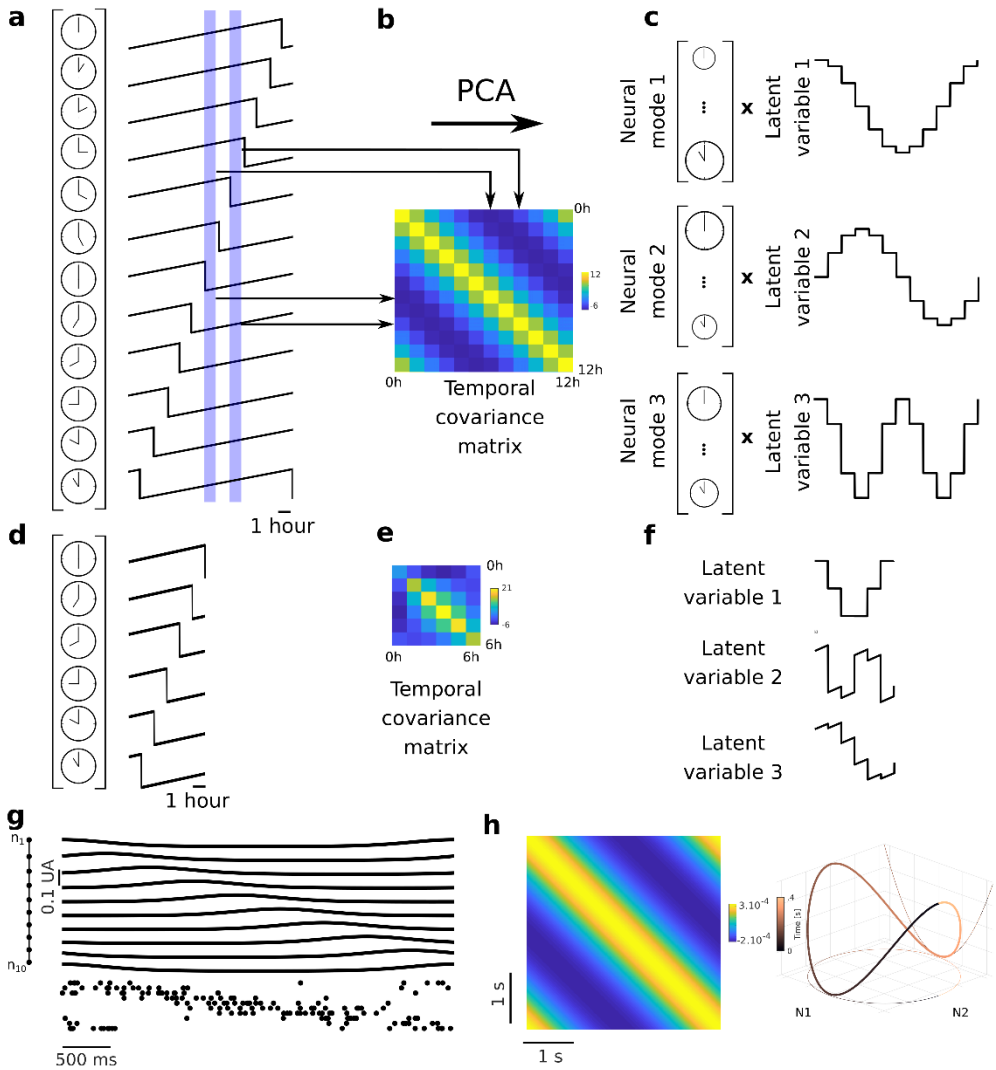
Ultimately, for this dataset, the latent variables will necessarily reproduce the Fourier series (cosines and sines ordered by increasing frequency, with the base frequency being given by one over the time length of the trial). The logic is as follows. The set – the mathematical definition of a collection of unordered elements – of all the measured absolute time differences $\tau=|t_2 - t1|$ collected at two given time instants t_1 and t_2 only depends on the time difference τ, modulo 24h, and not on the particular values of t_1 and t_2. We call this feature a "continuous variation in time". By systematically computing the difference between two set of watch measurements for all time pairs (t_i, t_j), we obtained the temporal covariance matrix. Because of the continuous variation in time, the temporal covariance matrix has a particular structure known as circulant matrix (Fig. 1b). It can be shown analytically (see Supplementary text) that the eigenvectors of this matrix, which are used to compute the latent variables after performing PCA, are given by the Fourier series[18]. Note that this will also be the case when using other dimensionality reduction methods, e.g. multidimensional scaling[17].

If we successively plot the latent variables of this watch dataset, we obtain sinusoidal curves of increasing frequency. These sinusoidal curves draw stereotypical patterns in two dimensions, known as Lissajous curves (as noted also by [15]). In three dimensions, it corresponds to the curve $[\cos(2\pi\, t/T), \sin(2\pi\, t/T), \cos(4\pi\, t/T)]$, with T the period given by one trial (Fig. 1c). These shapes represent the "horseshoe effect'[16], and closely resemble oscillatory trajectories often obtained when applying dimensionality reduction techniques to activity of neural populations. Yet, the oscillatory trajectories of these latent variables do not tell us if the set of watches generate latent oscillatory dynamics, they merely indicate the presence of a phase shift. For example, the time on each watch can be set by an external common source (e.g. earth rotation – an external oscillator with a circadian period - if all

watches were sundials: the watches themselves do not constitute a latent oscillatory dynamical system); by simply adding or subtracting time zone differences with respect to the one "calibration" (possibly external) watch (here again, watches are not a latent oscillatory dynamical system); or by intrinsically oscillating with a circadian period while interacting with other watches to adjust their time if needed (possible existence of latent oscillatory dynamical systems)[19].

Horseshoe effects are also observed if the measurements are not periodic. For instance, if we now select the measurements of six watches over a period of six hours, approximate sinusoidal curves that deviate from the exact Fourier series are observed (Fig. 1 d-f). These deviations occur because the temporal covariance matrix now has a Toeplitz structure[18]0020(Fig. 1e). Toeplitz matrix structure is similar to a circulant matrix structure, but without periodic boundary conditions, and the resulting latent variables can be analytically shown to be approximately sinusoidal curves[18].

The horseshoe effect can be similarly observed using simulations of neural spiking activity. As an example, we built a perfectly regular model by spatially arranging ten neurons on a line. The time evolution of the firing rate of each neuron is given by a periodic smooth function with a localized increase given by a phase shift that depends on the location of the neuron on the line (Fig. 1f, Methods). Note that no latent oscillatory dynamical systems are involved in generating this dataset. With this model, we obtain a circulant temporal covariance matrix and reproduce the Fourier series as predicted by the analytical derivation (Fig. 1g). When applying PCA to such signals, the oscillatory trajectories will always rotate counter-clockwise (see Supplementary text)[18]. This does not depend on the ordering of the neurons – shuffling them will produce the oscillatory trajectories that always rotate in the same direction.



[~ 250 words] ***Oscillatory trajectories emerge as latent variables after applying dimensionality reduction methods if the data exhibit continuous variation in time.*** *(**a**) 12 watches are used to record time in 12 different time zones, one watch for each zone. The dataset constitutes of the recorded times over 12 hours. (**b**) Computing the covariance between all pairs of time instant gives the temporal covariance matrix, which here has a circulant structure. (**c**) PCA applied on this dataset extracts neural modes and latent variables, which are sinusoidal curves of increasing frequencies. (**d**) We select 6 watches for 6 different time zones, with recordings spanning 6 hours. (**e**) Computing the covariance between all pairs of time instant gives the temporal covariance matrix, which here has an approximate Toeplitz structure. (**f**) Latent variables are approximately sinusoidal curves of increasing frequencies. (**g**) Simulated activity of 10 regularly sampled (over a total of 100) neurons using point process models whose intensity functions are given by cosine functions with distinct phase shifts. (**h**) Temporal covariance matrix is circulant (left panel), and the first three latent variables are sinusoidal curves of increasing frequency (right panel).*

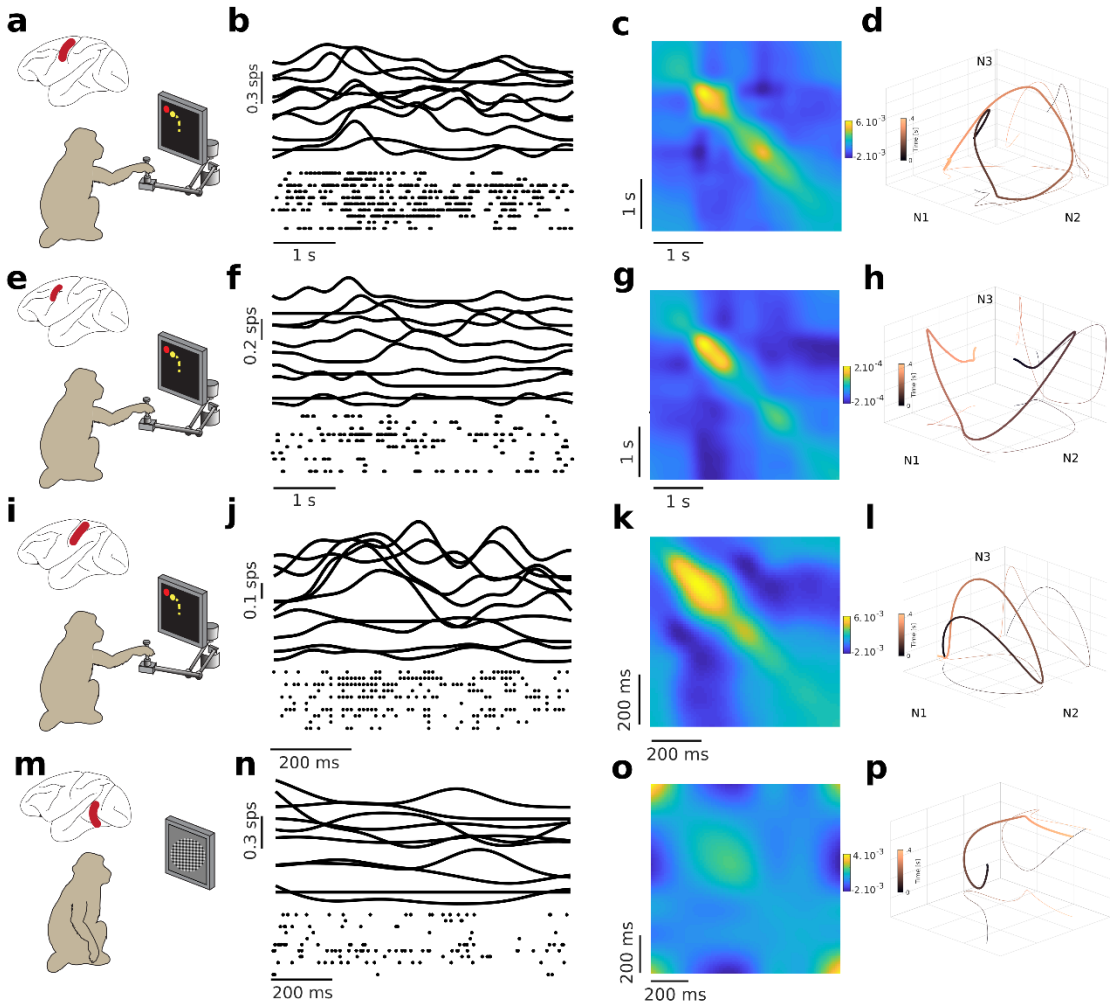
In many important aspects similar to our SfN meeting and spiking activity examples, microelectrode array recordings of the neuronal firing rates, which quasi-homogeneously sample neural population, should also in typical experimental scenarios approximately exhibit a continuous variation in time[19–24].

To show this, we considered several datasets of non-human primates performed various tasks. We start by analyzing the spiking activity of a rhesus macaque monkeys implanted with two 96-channel microelectrode arrays in the primary motor (M1) and premotor dorsal cortex (PMd) (Fig. 2a-b, e-f)[25]. The monkey is performing an instructed-delay center-out and reaching task, with eight reaching directions. The trial-averaged temporal covariance matrix for a single reaching direction resembled a circulant matrix (Fig. 2c, g) and, as expected according to the horseshoe effect, latent variables of neural activity recorded during reaches to the same direction displayed oscillatory trajectories that resembled sinusoidal curves of our SfN meeting example (Fig. 2d, h). Due to the horseshoe effect, the presence of oscillatory trajectories is not sufficient evidence to conclude that the neural activity was generated by latent oscillatory dynamical systems. Similar oscillatory trajectories were found when analyzing data from another rhesus macaque monkey implanted with a microelectrode array in the primary somatosensory cortex (S1) and performing the same task (Fig. 2i-l)[26,27].

We then looked at data recorded from a rhesus macaque monkey implanted with a microelectrode array in the V4 area[28,29] (Fig. 2m-p). The monkey was performing a fixation task while a sequence of static gratings and plaid stimuli were shown peripherally in a location centered on the V4 receptive fields. Each fixation included 10 flashed stimuli shown in a one second sequence for 100 ms each. We also found oscillatory trajectories in this dataset, although they deviated from the ideal model shown in Fig. 1h.

Consideration of the horseshoe effect described above may also help in the interpretation of several recent results, including long-term stability of recordings[9], measures of the dynamical character of a system such as tangling[4], and the partition of neural activity into distinct subspaces[30] (see Supplementary text for a detailed discussion).

Numerous studies[2–8,14,15] and the examples shown above described continuous variation in time (that is, sequence of firing patterns) as a generic property of neural activity that can be encountered in many brain areas, for very varied type of tasks, and for several types of recordings, suggesting that this property is a ubiquitous feature of neural activity. Reasons for which neural activity continuously varies in time is still debated, and could reflect a large variety of processes, including variable time offsets between neurons and dynamics[31], or sequence of firing patterns in neural activity[20]. Both Lebedev et al.[15] and Michaels et al.[14] have discussed the emerging horseshoe effect in M1 cortex, although they diverge on whether the data support or not the existence of a latent oscillatory dynamical system.



[~ 250 words] ***Horseshoe effect in neural data.*** *(**a**) Spiking activity was recorded from the motor cortex of a monkey performing a reaching task. (**b**) Top panel shows the firing activity for 10 example neurons, and bottom panel shows the corresponding spiking activity. (**c**) Temporal covariance matrix of all 84 neurons approximate a Toeplitz matrix. (**d**) Neural activity projected in a space spanned by the three leading modes (N1, N2 and N3) traces three nearly sinusoidal curves. (**e-h**) Similar plots to (a-d) for the same monkey and the same task showing recordings from the premotor cortex instead. (**i-l**) Similar plots to (a-d) for another monkey and the same task showing recordings from the primary sensory cortex. (**m-p**) Similar plots to (a-d) recorded from the area V4 of a monkey performing a fixation task while visual stimuli were shown. Monkey images adapted from SciDraw.io.*

Oscillatory trajectories obtained after dimensionality reduction of neural activity deviate from the exact Fourier series predicted by the ideal model (e.g. the SfN meeting example in Fig. 1a-c, or neural model in Fig. 1g-h). These deviations occur because the temporal sequence of firing patterns do not vary perfectly homogeneously in time. In particular, sequence variations and therefore temporal covariance matrices are task and conditions dependent. For instance, in the center-out monkey dataset (Fig. 2a-d), temporal covariance

matrices slightly differ for each reaching direction. This is evidenced by applying PCA on the same center-out monkey dataset, but concatenating trials for two opposite reaching conditions. This resulted in latent variables tracing two quasi-orthogonal oscillatory trajectories (Fig. 3a). Applying PCA on the full center-out monkey dataset containing 8 reaching conditions, resulted in three leading latent variables tracing eight oscillatory trajectories arranged by the reaching angle (Fig. 3b). The temporal covariance matrix of this dataset has a block structure, each block corresponding to a direction, and blocks relative to neighboring reaching directions more similar to each other's than blocks for more distant reaching directions. The ideal model presented above does not account for these deviations. In the following, we show that modeling these deviations bring useful information about the dimensionality and structure of the encoding space.

We quantified how much recorded signals diverge from our simulations in two ways. First, we regressed the values of the temporal covariance matrix calculated from the recorded data to the temporal covariance matrix of a model using Pearson's correlation measure R^2. Second, we quantified the match between latent variables and sinusoidal curves by applying canonical correlation analysis (CCA) and computing the canonical correlation (CC) between those two trajectories. CC evaluates how well two multidimensional trajectories correlate after linear transformations that optimize their matching. For instance, for the experiments shown in Fig. 2, correlations between recorded data and the ideal model were as follows: R^2=0.74 and CC=0.92 for the M1 data (Fig. 2a-d), R^2=0.65 and CC=0.92 for the PMd data (Fig. 2e-h), R^2=0.66 and CC=0.83 for the S1 data (Fig. 2i-l), R^2=0.55 and CC=0.80 for the V4 data (Fig. 2m-p). These measures can also be used to systematically evaluate the impact of biased sampling on the temporal covariance matrices and resulting trajectories (Supplementary Fig. 1 and 2).

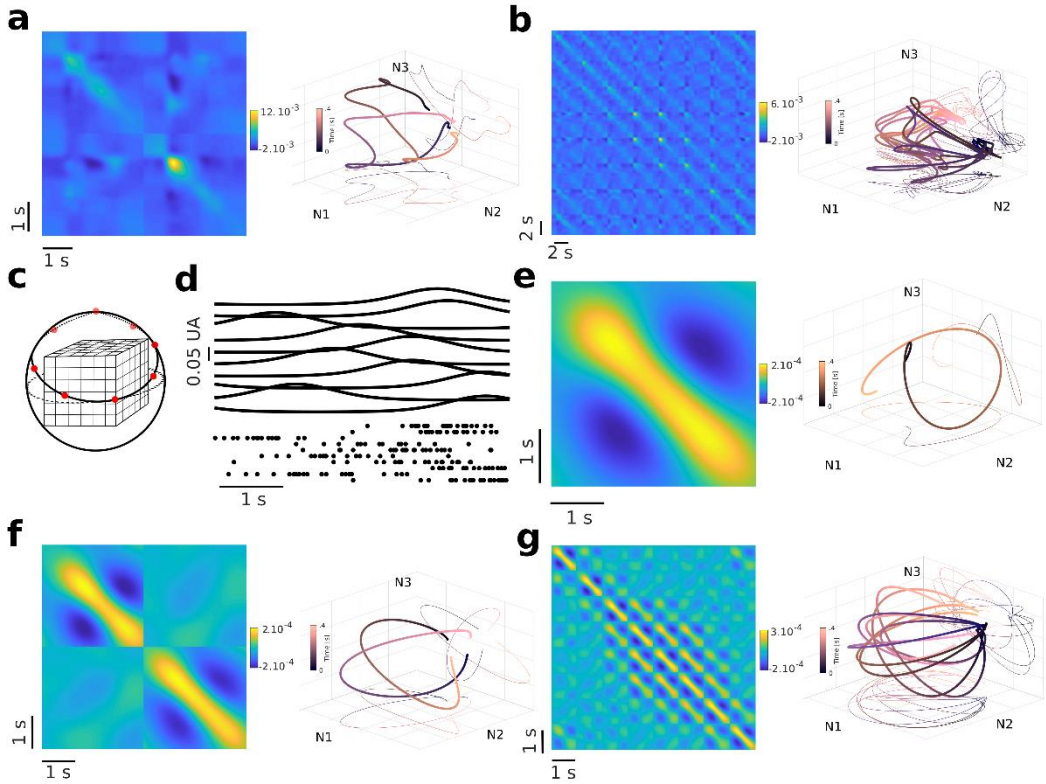
To reproduce the deviations of the oscillatory trajectories observed in the recordings for the eight reaching directions, we simulate a cubic volume of 216 neurons where time shift between neural activation of neurons is directly encoded as a spatial variable (Fig. 3c). Intuitively, we create a 6x6x6 three-dimensional functional space where synaptic connections between neurons are encoded by their physical proximity. This does not entail that neurons are physically organized in three dimensional space, but rather than the connections between neurons can be unfolded in a three-dimensional space where distance between neurons encore their functional similarity. The firing rate of neurons is given by a periodic smooth function, more or less shifted depending on the location of the neuron in the cube. Each reaching direction dictates the value of this shift for each neuron, which is computed as the distance from each neuron to a source whose location outside the cube correspond to the reaching direction. In practice, our model simulations resemble a wave propagating from each source across the cube in a different direction (Supplementary Fig. 3, Methods). Note that while we have chosen to construct this model as representational without latent oscillatory dynamical systems involved in generating the datasets, it can equally be described by a dynamical system that models a spatio-temporal wave propagation with a distinct source for each encoded direction. For model comparison, we also created alternative one and two-dimensional models (Supplementary Fig. 3, Methods).

We first used our model to simulate one reaching direction in the center-out task (Fig. 3d-e). Temporal covariance matrix and the first three neural modes closely resemble the data for

one reaching direction (Fig. 2c-d). Correlations between recorded data shown in Fig. 1d and our model were high (R^2=0.76, CC=0.94) and were slightly better than alternative models (R^2=0.67, CC=0.91 for one-dimensional model, R^2=0.72, CC=0.94 for two-dimensional model, Supp. Fig. 4).

We then simulated two opposite sources, one for each reaching directions (Fig. 3f). When we perform PCA on this simulated dataset, we also obtained quasi-orthogonal sinusoidal curves that resemble the data (Fig. 3a). Temporal covariance matrix of motor cortical neural activity was a block matrix, each block corresponding to a direction. The three dimensional model performed noticeably better (R^2=0.35, CC=0.80) than the one- and two- dimensional models (R^2=0.22, CC=0.70 for one dimensional model, R^2=0.37, CC=0.70 for two dimensional model, Supp. Fig. 4).

We finally used the same model to simulate 8 reaching direction (Fig. 3g). The origins of the 8 directions and dimension of the functional space we used for the model can be deduced from observing the temporal covariance matrix obtained from the data (Fig. 3b): close-by directions lead to similar neural activity in a circular fashion, yet neural activity for opposing direction are close to orthogonal, warranting a 3d functional space (Supplementary Fig. 3). We applied PCA on this data and obtained oscillatory trajectories in the first three neural modes that resemble the recorded 8-target dataset (CC=0.28, Fig. 3g). This model performed better than the one-dimensional and two-dimensional models (CC=0.03 and CC=0.14 respectively, Supp. Fig. 4). Similarity of the dataset properties was confirmed by the similarity between temporal covariance matrices (R^2=0.64, Fig. 3g), and was also better than when using alternative models (R^2=0.35 for one dimensional model, R^2=0.46 for two dimensional model). Thus, by identifying the deviations between the trajectories associated with each reaching direction, our model provides a convenient tool to help identifying the necessary structure and the dimensionality of the space encoding the functional proximity between neurons.



[~ 250 words] ***Deviations of oscillatory trajectories for multiple conditions inform models of neural activity****. (**a**) Colorplot in the left panel shows the temporal covariance matrix for 2 reaching directions. Neural activity projected in the space spanned by three leading modes traces 2 oscillatory trajectories, one for each direction (right panel). (**b**) Colorplot in the left panel shows the temporal covariance matrix for 8 reaching directions. Neural activity projected in the space spanned by three leading modes traces 8 oscillatory trajectories, one for each direction (right panel). (**c**) We designed a neural encoding model that lacks latent oscillatory dynamical systems to demonstrate that it can accurately reproduce deviations across movements from the ideal circulant matrix. For each trial representing a movement in one of eight directions, we simulated the functional proximity of activity in neurons as a cubic spatial ordering with the phase shift of each neuron given by its distance each of the 8 sources for each reaching direction respectively. Red points show the sources. (**d**) Simulated firing rate (top panel) and corresponding spiking activity (bottom panels) can reproduce similar dynamics. (**e**) Temporal covariance matrix is nearly Toeplitz. The first three latent variables (right panel) are nearly sinusoidal curves of increasing frequencies (**f**) Temporal covariance matrix for 2 simulated reaching directions resembles the temporal covariance matrix of the real data (left panel). Neural activity projected in the space spanned by three leading modes traces 2 oscillatory trajectories similar to those observed in real data (right panel). (**g**) Temporal covariance matrix for 8 simulated trials also resembles the temporal covariance matrix of the real data (left panel). Neural activity projected in the space spanned by three leading modes traces 8 oscillatory trajectories similar to those observed in real data (right panel).*

In this article, we showed that interpreting the output of dimensionality reductions techniques applied to neural data can be sometimes misleading. Despite their limitations,

dimensionality reduction techniques are extremely useful for summarizing the increasingly complex high-dimensional neural recordings. We proposed a new simple, intuitive and tractable model that goes beyond the dichotomy between representational and dynamical models to focus on the deviations between conditions and directly encode time shift between neurons as a spatial variable. This model brings insights about the encoding of the tasks in neuronal networks by identifying latent structures in the temporal correlation of the neural activity. We expect that the dimensionality of the model will need to increase with the complexity of the task[32,33], which might help bringing a testable account of generalization to natural behavior[34].

References

1. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
2. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of Movement. *Neuron* **94**, 978–984 (2017).
3. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* (2012) doi:10.1038/nature11129.
4. Russo, A. A. *et al.* Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response. *Neuron* **97**, 953-966.e8 (2018).
5. Rutten, V. M. S., Bernacchia, A., Sahani, M. & Hennequin, G. Non-reversible Gaussian processes for identifying latent dynamical structure in neural data. 21.
6. Trautmann, E. M. *et al.* Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron* **103**, 292-308.e4 (2019).
7. Kobak, D. *et al.* Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
8. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
9. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
10. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* **40**, 646–649 (2008).
11. Podani, J. & Miklós, I. Resemblance Coefficients and the Horseshoe Effect in Principal Coordinates Analysis. *Ecology* **83**, 3331–3343 (2002).
12. Richman, M. B. Rotation of principal components. *J. Climatol.* **6**, 293–335 (1986).
13. Brillinger, D. R. *Time series: data analysis and theory*. (Holden-Day, 1981).
14. Michaels, J. A., Dann, B. & Scherberger, H. Neural Population Dynamics during Reaching Are Better Explained by a Dynamical System than Representational Tuning. *PLOS Comput. Biol.* **12**, e1005175 (2016).

15. Lebedev, M. A. *et al.* Analysis of neuronal ensemble activity reveals the pitfalls and shortcomings of rotation dynamics. *Sci. Rep.* **9**, (2019).
16. Kendall, D. G. A mathematical approach to seriation. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **269**, 125–135 (1970).
17. Diaconis, P., Goel, S. & Holmes, S. Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat.* **2**, 777–807 (2008).
18. Gray, R. M. Toeplitz and Circulant Matrices: A Review. *Found. Trends® Commun. Inf. Theory* **2**, 155–239 (2006).
19. Ermentrout, G. B. & Kleinfeld, D. Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron* **29**, 33–44 (2001).
20. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
21. Luczak, A., Barthó, P., Marguet, S. L., Buzsáki, G. & Harris, K. D. Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl. Acad. Sci.* **104**, 347–352 (2007).
22. Bulkin, D. A., Law, L. M. & Smith, D. M. Placing memories in context: Hippocampal representations promote retrieval of appropriate memories. *Hippocampus* **26**, 958–971 (2016).
23. Kvitsiani, D. *et al.* Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).
24. Gage, G. J., Stoetzner, C. R., Wiltschko, A. B. & Berke, J. D. Selective Activation of Striatal Fast-Spiking Interneurons during Choice Execution. *Neuron* **67**, 466–479 (2010).
25. Perich, M. G., Gallego, J. A. & Miller, L. E. A Neural Population Mechanism for Rapid Learning. *Neuron* **100**, 964-976.e7 (2018).
26. Chowdhury, R. H., Glaser, J. I. & Miller, L. E. Area 2 of primary somatosensory cortex encodes kinematics of the whole arm. *eLife* **9**, e48198 (2020).
27. Chowdhury, R., Glaser, J. & Miller, L. Data from: Area 2 of primary somatosensory cortex encodes kinematics of the whole arm. 4016198198 bytes (2020) doi:10.5061/DRYAD.NK98SF7Q7.
28. Snyder, A. C., Yu, B. M. & Smith, M. A. Distinct population codes for attention in the absence and presence of visual stimulation. *Nat. Commun.* **9**, 4382 (2018).
29. Smith, M. *V4 Utah Array Plaid Movie Data. figshare. Dataset.* (2020).
30. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).
31. Moran, D. W. & Schwartz, A. B. Motor Cortical Representation of Speed and Direction During Reaching. *J. Neurophysiol.* **82**, 2676–2692 (1999).
32. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* **70**, 113–120 (2021).
33. Humphries, M. D. Strong and weak principles of neural dimension reduction. *ArXiv201108088 Q-Bio* (2021).

34. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* 1–37 (2020) doi:10.1017/S0140525X20001685.

Acknowledgments

We thank Raeed H Chowdhury and Lee E. Miller for collecting the S1 dataset. We thank Matthew Smith for collecting the V4 dataset. TP is supported by the Swiss National Science Foundation (Grant #193542). The study was supported in part by the Swiss National Science Foundation Ambizione grant (PZ00P2_168103).

Author contributions

T.P. and T.M. designed the study. M.G.P collected the M1 and PMd datasets. T.P. performed the analysis and modeling. T.P., M.G.P. and T.M. wrote the manuscript.

Competing financial interests

The authors declare no competing interests.

Methods

All recording methods used in this article have been described in previous publications[25,26,28]. We here only briefly describe them.

Experimental subjects

The animal neural datasets presented in this article were collected in accordance with the ethical standards of the Northwestern University and the University of Pittsburgh Institutional Animal Care and Use Committee. Data collection procedures were consistent with Federal guidelines. The neural data were recorded from three male mucaca mulatta monkeys.

Behavioral tasks

For the M1, PMd, and S1 datasets, monkeys set in a non-human primate chair and controlled a cursor displayed on a computer screen using a custom 2-D planar manipulandum. To successfully end a trial, the monkeys performed a standard center-out task – they started from the center and captured one of eight different 2 x 2 cm square targets that were evenly distributed on a circle of an 8 cm radius. We recorded the manipulandum handle position at a 1k-sample frequency. The behavioral task was controlled by a custom Simulink code (The Mathworks, Inc). The trial started when the monkeys moved the manipulandum to the

center and captured the center target. Following a variable hold period (0.5 – 1.5 s), one of the eight outer targets appeared. After a variable instructed delay period (0.5 – 1.5 s), the monkeys received an auditory "go" cue, and the center target disappeared. The monkeys had to reach the outer target within one second, and had to hold the cursor there for another 0.5 s to successfully finish the trial. For the V4 dataset, monkeys performed a fixation task while a sequence of static gratings and plaid stimuli were shown peripherally in a location centered on the V4 receptive fields. Each trial started with a period of fixation, followed by a sequence of ten flashed stimuli shown for 100 ms each and a new fixation period. The trial ended with a saccade to a random position target.

Data acquisition

Monkeys were surgically implanted with microelectrode arrays (Blackrock Microsystems, Salt Lake City, UT), simultaneously in M1 and PMd for the first monkey, simultaneously in M1 and S1 for the second monkey, and in V4 for the third monkey. Blackrock Cerebus system (Blackrock Microsystems, Salt Lake City, UT) were used to record 96 channels of neural activity from each array. For the M1, PMd, and S1 datasets, the acquired data were processed using the Offline Sorter v3 software (Plexon, Inc, Dallas, TX) to identify action potential trains of individual neurons. There is a small possibility that duplicate neural activity can appear on different channels due to electrode shunting. We performed two procedures to ensure our data included only independent channels. First, we used Blackrock Cerebus system utility to identify and disable any channels with high crosstalk. Second, we computed the percent of coincident spikes between any two channels. This percentage was compared against an empirical probability distribution from all sessions of data. Any neurons with a coincidence above a 95% probability threshold were excluded from the analysis (approximately 15%–20% neurons per session). For the V4 dataset, data were sorted using customized semi-supervised clustering. Neurons that were not recorded stably throughout the session were excluded. For all datasets, we excluded neurons with trial-averaged firing rates below 1 Hz.

Data processing

Neural spiking activity was counted in 10 ms bins. A smooth firing rate was then obtained for each electrodes by convolving a Gaussian kernel (s.d. 100 ms) with the binned spike train. For each trial of the center-out reaching task, we considered all time points between the go cue and the return to the initial start position after successful reaching of the target. Trials that were outside of the range of the median trial duration plus or minus 200 ms were discarded. To ensure that each trial contains the same number of samples (which is required for computing canonical correlation), we linearly interpolated reaching and returning sequence of each trial to match the number of samples found in the median trial duration. For the fixation task with visual stimuli, we used the full one-second length of the trials, shifted by 50 ms to account for the latency between the visual signal and the observed activity in V4. We used 124 trials of the same repeated stimulus sequence.

Models

SfN watch model

We simulated the time on a set of 12 watches. The time on each watch linearly increases from 0 h to 12 h, and is reset to 0 h whenever it reaches 12 h. For each of the watches, the initial conditions were given by a phase shift specific to each watch, and ranged from 0 h to 11 h, e.g. time on watch 1 start at 0 h, time on watch 2 starts at 1 h, etc.

Ideal neuronal model

We simulated 216 neurons using a point-process generalized linear model with a log-link function and a conditionally Poisson distribution. The dynamics of the instantaneous rate (also called conditional intensity function) of neuron i were modeled as:

$$log(\lambda_i(t|\varphi_i)\Delta t) = \mu + acos(2\pi t + \varphi_i) \qquad (1)$$

with the phase $\varphi_i \in \left(0, \frac{2\pi}{n}, \dots, 2\pi\right)$, $\mu = -5$, and $a = 2$. In this and all subsequent simulations, the number of time samples were chosen to match the number of samples of the data (396 for the M1 and PMd datasets, 74 for the S1 dataset, and 100 for the V4 dataset).

Full neuronal model

We simulated 216 neurons. The dynamics of the instantaneous rate of neuron i were modeled by Eq. (1) with $\mu = -4$, and $a = 2$. The phase φ_i is calculated as the distance between the position of each neuron in a three dimensional space and the position of the source, which is different for each reaching direction:

$$\varphi_i = 2\pi \, \|x^T - s^T\|_2 / \max_i \left(\|x^T - s^T\|_2\right) \qquad (2)$$

with each neuron i and source j given as a set of three Cartesian coordinates $x = (x_i, y_i, z_i)$ and $s(s| \quad |j^x, s_j^y, s_j^z)$ (Supplementary Fig. 3). For neurons, $x_i \in [\![1,6]\!]$, $y_i \in [\![1,6]\!]$, and $z_i \in [\![1,6]\!]$. For sources s_j, $j \in [\![1,8]\!]$, we used two spherical coordinates $(\theta| \quad |j, \phi_j) \in \left(\left(0, \frac{\pi}{2}\right), \left(\frac{\pi}{4}, \frac{3\pi}{8}\right), \left(\frac{\pi}{2}, \frac{\pi}{4}\right), \left(\frac{3\pi}{4}, \frac{\pi}{8}\right), (\pi, 0), \left(\frac{5\pi}{4}, \frac{\pi}{8}\right), \left(\frac{3\pi}{2}, \frac{\pi}{4}\right), \left(\frac{7\pi}{4}, \frac{3\pi}{8}\right)\right)$ and a radius $r = 2$.

Alternative one-dimensional model

We simulated 217 neurons. The dynamics of the instantaneous rate of neuron i were modeled by Eq. (1) with $\mu = -4$, and $a = 2$, and the phase as for the full neuronal model defined in Eq. (2) (Supplementary Fig. 3). Each neuron i and source j were given as a one Cartesian coordinates $x = (x_i)$ and s_j^x. For neurons, $x_i \in [\![1,217]\!]$. Sources s_j, $j \in [\![1,8]\!]$, were equally spread over the line of neurons: $s_j^x = (j-1) * 31 + 1$.

Alternative two-dimensional model

We simulated 225 neurons. The dynamics of the instantaneous rate of neuron i were modeled by Eq. (1) with $\mu = -4$, and $a = 2$, and the phase as for the full neuronal model defined in Eq. (2) (Supplementary Fig. 3). Each neuron i and source j were given as a set of two Cartesian coordinates $x = (x_i, y_i)$ and $\left(s\middle| \quad \middle|j^x, s_j^y\right)$. For neurons, $x_i \in [\![1,15]\!]$, and $y_i \in [\![1,15]\!]$. For sources s_j, $j \in [\![1,8]\!]$, we used one polar coordinate $\theta_j \in \left(0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}, \pi\right)$ and a radius $r = 10.61$.

Dimensionality reduction

Principal component analysis

We used standard principal component analysis (PCA). PCA projects the firing rate, originally described in the basis built by using each neuron as a dimension, to a new orthonormal basis, by maximizing the variance of the projected data[1]. We termed this new basis the neural modes, and the variables dictating the evolution of these neural modes in time are called latent variables, following previous taxonomy[9] . Formally, if we defined our feature matrix $X = \begin{bmatrix} x_1^1 & \cdots & x_1^N \\ \vdots & \ddots & \vdots \\ x_t^1 & \cdots & x_t^N \end{bmatrix}$, with t the length of the trials and N the number of neurons, then the neural modes and latent variables are given by the eigenvectors and eigenvalues of the covariance matrix $E\left[\left(X - \overline{X}\right)\left(X - \overline{X}\right)^T\right]$. To compute the temporal covariance matrix, we compute the covariance between the transposed feature matrices, $E\left[\left(X - \overline{X}\right)^T\left(X - \overline{X}\right)\right]$.

Measures

Canonical correlation

Canonical correlation finds linear transformations between two distinct feature matrices such that the corresponding latent dynamics are maximally correlated[9]. Formally, this corresponds to recursively defining a new basis (ξ_X, ξ_Y) for each feature matrix X and Y such that each of the dimension $i \in [\![1, N]\!]$ of the basis maximizes the canonical correlation $\mathrm{corr}\left(\xi_X^i X, \xi_Y^i Y\right)$ and is orthogonal to previously defined components.

Correlation between covariance matrices

Correlation between two temporal covariance matrices was assessed by using the coefficient of determination between all the values of those covariance matrices. It was obtained by linearly regressing all values of one temporal covariance matrix on all values of the other temporal covariance matrix.

Data availability

M1 and PMd data were used in previously published article[25] and are available upon requests to the authors of this publication. S1 data are available on Dryad[26]. V4 data are available on CRCNS.org[29].

Code availability

Code needed to run the data processing, analysis and modeling is available at **DOI will be made available upon acceptance**.