REVIEW

Neural manifold analysis of brain circuit dynamics in health and disease

Rufus Mitchell-Heggs[1,2] · Seigfred Prado[1,3] · Giuseppe P. Gava[1] · Mary Ann Go[1] · Simon R. Schultz[1]

Abstract

Recent developments in experimental neuroscience make it possible to simultaneously record the activity of thousands of neurons. However, the development of analysis approaches for such large-scale neural recordings have been slower than those applicable to single-cell experiments. One approach that has gained recent popularity is neural manifold learning. This approach takes advantage of the fact that often, even though neural datasets may be very high dimensional, the dynamics of neural activity tends to traverse a much lower-dimensional space. The topological structures formed by these low-dimensional neural subspaces are referred to as "neural manifolds", and may potentially provide insight linking neural circuit dynamics with cognitive function and behavioral performance. In this paper we review a number of linear and non-linear approaches to neural manifold learning, including principal component analysis (PCA), multi-dimensional scaling (MDS), Isomap, locally linear embedding (LLE), Laplacian eigenmaps (LEM), t-SNE, and uniform manifold approximation and projection (UMAP). We outline these methods under a common mathematical nomenclature, and compare their advantages and disadvantages with respect to their use for neural data analysis. We apply them to a number of datasets from published literature, comparing the manifolds that result from their application to hippocampal place cells, motor cortical neurons during a reaching task, and prefrontal cortical neurons during a multi-behavior task. We find that in many circumstances linear algorithms produce similar results to non-linear methods, although in particular cases where the behavioral complexity is greater, non-linear methods tend to find lower-dimensional manifolds, at the possible expense of interpretability. We demonstrate that these methods are applicable to the study of neurological disorders through simulation of a mouse model of Alzheimer's Disease, and speculate that neural manifold analysis may help us to understand the circuit-level consequences of molecular and cellular neuropathology.

✉ Simon R. Schultz
s.schultz@imperial.ac.uk

Rufus Mitchell-Heggs
rmitch11@ed.ac.uk

[1] Department of Bioengineering and Centre for Neurotechnology, Imperial College London, London SW7 2AZ, United Kingdom

[2] Centre for Discovery Brain Sciences, The University of Edinburgh, Edinburgh EH8 9XD, United Kingdom

[3] Department of Electronics Engineering, University of Santo Tomas, Manila, Philippines

1 Introduction

While the investigation of single neurons has undoubtedly told us much about brain function, it is uncertain whether individual neuron properties alone are sufficient for understanding the neurobiological basis of behavior (Pang et al., 2016). In some cases, trial-averaging of single-neuron responses may lead to confusing or misleading interpretation of true biological mechanisms (Sanger & Kalaska, 2014; Cunningham & Yu, 2014). Additionally, single-neuron activities studied in higher-level brain areas involved in cognitive tasks (Machens et al., 2010; Laurent, 2002; Churchland et al., 2010) are highly heterogeneous both across neurons and across experimental conditions even for nominally identical trials. And finally, it may well be that task-relevant

information is represented in patterns of activity across multiple neurons, above and beyond what is observable at the single neuron level. Unfortunately, characterizing such patterns, in the worst case, may require measurement of an exponential number of parameters (the "curse of dimensionality").

However, it appears that in many circumstances, patterns of neural population activity may be described in terms of far fewer population-level features than either the number of possible patterns, or even the number of neurons observed (Churchland et al., 2012; Mante et al., 2013; Chaudhuri et al., 2019; Gallego et al., 2020; Nieh et al., 2021; Stringer et al., 2019). This underpins a paradigm shift in the studies of neural systems from single-neuron analysis to a more comprehensive analysis that integrates single-neuron models with neural population analysis.

If, as appears to be the case, the spatiotemporal dynamics of brain activity is low dimensional, or at least much lower-dimensional than the pattern space, then it stands to reason that such activity can be characterized without falling afoul of the curse of dimensionality, and on reasonable experimental timescales. Indeed, in recent years, numerous techniques have been developed to do just that. For instance, classification algorithms have been applied to neuronal ensembles to predict aspects of behavior (Rigotti et al., 2013; Fusi et al., 2016; Rust, 2014; Raposo et al., 2014). One problem with this is that the common practice is to identify "neuronal ensembles" by grouping together neurons with sufficiently highly correlated activity during the same behaviors or in response to the same stimuli. This ignores information that is transmitted collectively and might lead to (i) falsely concluding that a group of neurons do not encode a behavioral variable (when in fact they encode it collectively), (ii) incorrectly estimating the amount of information that is being encoded, and/or (iii) missing important mechanisms that contribute to encoding (Frost et al., 2021).

Alternatively, artificial neural networks (ANN) have been increasingly employed, either by (i) using a goal-driven neural network and using the embedding to compare and predict the population activity in different brain regions (Russo et al., 2018; Mante et al., 2013; Jazayeri and Ostojic, 2021), or (ii) modelling the activity of individual neurons as emanating from a feedforward or recurrent neural network architecture (Elsayed et al., 2016; Rajan et al., 2016). Whilst these methods can present powerful ways of inferring neural states and dynamics, some issues have been raised on their biological interpretability, even though recent work has addressed some of them, as we discuss in Section 2.11.

To address these shortcomings, a range of techniques which are commonly referred to under the umbrella term of "neural manifold learning" (NML) have been employed. Some of these approaches simply make use of long-established general methods for dimensionality reduction (such as principal components analysis), whereas others have been developed specifically to study high-dimensional neural datasets.

Mathematically speaking, a manifold is a topological space that locally resembles our usual Euclidean space. If we form a multivariate time-series by convolving the spike trains of a neural population with a smoothing filter, and consider the activity pattern across these time series at each time to occupy a point in a neural activity space, then over time the activity will excurse a subspace that has often been observed to appear like such a manifold. Characterizing the geometry of such structures may offer important insights into neural computation (Chung & Abbott, 2021). In practical terms, the "neural manifold" is a low-dimensional subspace within the higher-dimensional space of neural activity which explains the majority of the variance of the neural dynamics (Fig. 1). Of course, real neural population dynamics are subject to noise, and in real experiments the topological subspace that can be excursed by the dynamics of neural activity can only be sampled, often sparsely. We
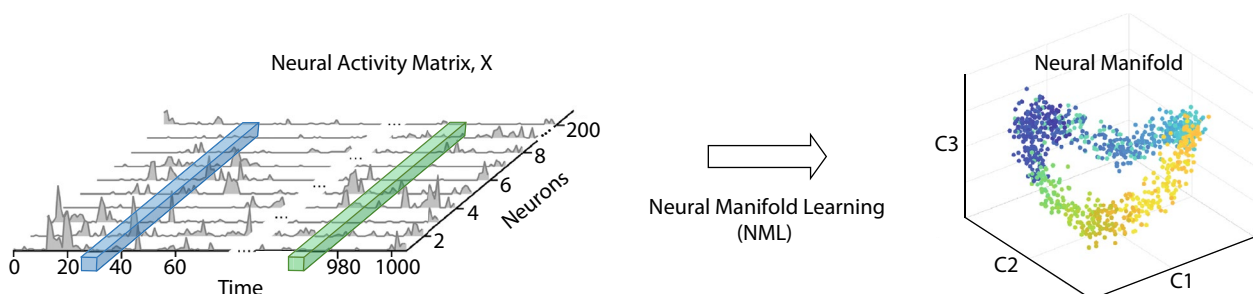

Fig. 1 Schematic showing a typical example of how a manifold learning algorithm may reduce the dimensionality of a high-dimensional neural population time series to produce a more interpretable low-dimensional representation. A high-dimensional neural population activity matrix, $\mathbf{X}$, with N neurons and T time points, is projected into a lower-dimensional manifold space and the trajectory visualized in the space formed by the first 3 dimensions, $c1$, $c2$ and $c3$

must make several important comments here: firstly, that this characterization of neural systems depends inherently upon a description of the system in which the state is continuous and determined by the instantaneous firing rates of each neuron (although those firing rates might be calculated from filters implementing shorter or longer time windows). And secondly, that what may be observed experimentally is typically a "point cloud" - however these points do not themselves constitute a manifold; instead they are normally taken as indicative of the underlying topological space which they sample. And finally, it should be noted that the term "neural manifold" is frequently used relatively loosely in neuroscience to refer to any kind of low-dimensional structure which may or may not meet the criteria of a mathematical manifold. In this survey article we will not dwell overly on this distinction, noting it but considering that utility is at this stage more important, and that it is likely that terminology will continue to evolve.

Neural manifold learning algorithms are algorithms for efficiently extracting a description of such a subspace from a sample of multivariate neural activity. Here we review how neural manifold learning can be employed to extract low-dimensional descriptions of the internal structure of ensemble activity patterns, and used to obtain insight into how interconnected populations of neurons represent and process information. Such techniques have been applied to neural population activity in a variety of different animals, brain regions and during distinct network states. In this review we compare a variety of neural manifold learning algorithms with several datasets from multi-electrode array electrophysiology and multi-photon calcium imaging experiments, to assess their relative merits in gaining insight into the recorded neural dynamics and the computations they may be underpinning.

We envisage that neural manifold learning may not only facilitate accurate quantitative characterizations of neural dynamics in healthy states, but also in complex nervous system disorders. Anatomical and functional brain networks may be constructed and analyzed from neuroimaging data in describing and predicting the clinical syndromes that result from neuropathology. NML can offer theoretical insight into progressive neurodegeneration, neuropsychological dysfunction, and potential anatomical targets for different therapeutic interventions. For example, investigating neural populations in the medial prefrontal cortex that are active during social situations while encoding social decisions may enable hypothesis testing for disorders such as Autistic Spectrum Disorder (ASD) or Schizophrenia (Irimia et al., 2018; Kingsbury et al., 2019).

In this review, we introduce NML as a methodology for neural population analysis and showcase its application to the analysis of different types of neural data with differing behavioral complexity both in healthy and disease model states. For selected algorithms, we visualize how they represent neural activity in lower-dimensional embeddings and evaluate them on their ability to discriminate between neural states and reconstruct neural data. We aim to offer the reader the prospect of selecting with confidence the type of NML method that works best for a particular type of neural data, as well as an appreciation of how NML can be leveraged as a powerful tool in deciphering more precisely the basis of different cognitive impairments and brain disorders.

2 Neural manifold learning

Neural manifold learning (NML) describes a subset of machine learning algorithms that take a high-dimensional neural activity matrix $\mathbf{X}$ comprised of the activity of N neurons at T time points and embed it into a lower-dimensional matrix $\mathbf{Y}$ while preserving some aspects of the information content of the original matrix $\mathbf{X}$ - e.g. mapping nearby points in the neural activity space $\mathbf{X}$ to nearby points in $\mathbf{Y}$ (see Fig. 1) (Cunningham & Yu, 2014; Churchland et al., 2012; Meshulam et al., 2017; Mante et al., 2013; Harvey et al., 2012; Wu et al., 2017). When projected into this lower-dimensional space, the set of neural activity patterns *observed* are typically constrained within a topological structure, or manifold, $\mathcal{Y}$, which might have a globally curved geometry but a locally linear one. For instance, if the reduced dimensionality embedding matrix $\mathbf{Y}$ is three-dimensional (3D) (often depicted on a two-dimensional (2D) page for convenience of illustration), the neural manifold $\mathcal{Y}$ might describe a closed surface within that 3D space. Another way of looking at this is that the data points in $\mathbf{X}$ lie on a lower-dimensional manifold that can be parameterised by a lower-dimensional coordinate system give by $\mathbf{Y}$, and that the task of the manifold learning algorithm is to find that coordinate system. This approach has found widespread recent use across neuroscientific studies (Fig. 2), including for understanding neural mechanisms during speech (Bouchard et al., 2013), decision-making in prefrontal cortex (Mante et al., 2013; Harvey et al., 2012; Briggman et al., 2005; Stokes et al., 2013), movement preparation and execution in the motor cortices (Churchland et al., 2012; Kaufman et al., 2014; Yu et al., 2009; Feulner & Clopath, 2021; Gallego et al., 2017) and spatial navigation systems (Chaudhuri et al., 2019; Nieh et al., 2021; Rubin et al., 2019; Gardner et al., 2022).

2.1 Manifold learning algorithms

There are several types of manifold learning algorithms that can generally be divided into linear and non-linear approaches. Although they have similar goals, they may differ in the way they transform the data and in the type of
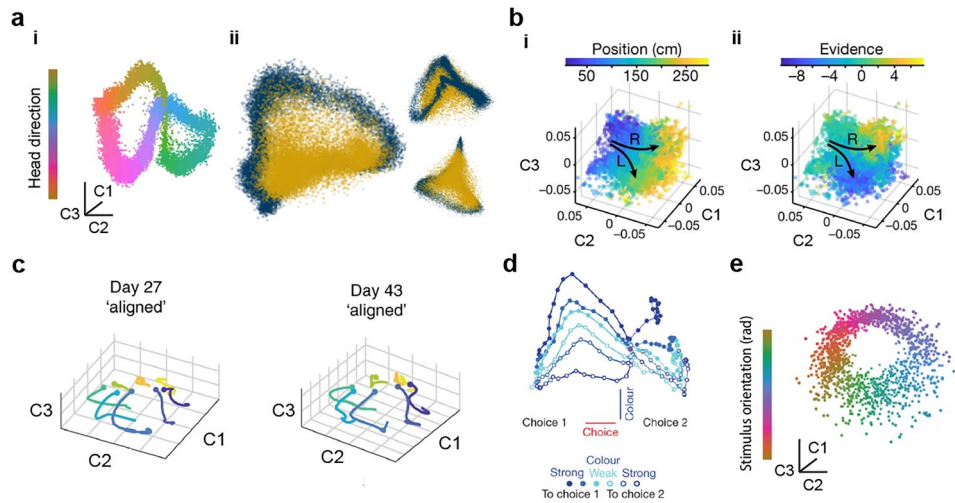
Fig. 2 Neural manifolds across different species and brain regions. **a** Population activity in the mouse head-direction circuit (Chaudhuri et al., 2019). **i** During waking, the network activity directly maps onto the head orientation of the animal. **ii** Comparison between population activity during waking (dark blue) and nREM sleep (mustard yellow); the latter does not follow the same one-dimensional waking dynamics. **b** Population activity in the mouse hippocampus during an evidence accumulation and decision making task in virtual reality (Nieh et al., 2021). Task-relevant variables such as **i** position and **ii** accumulated evidence are encoded in the manifold. **c** The motor cortical

population activity underpinning a reaching task in monkeys is stable over days and years (Gallego et al., 2020). **d** Prefrontal cortical population activity in macaque monkeys during a visual discrimination task spans a low-dimensional space. Task-relevant variables such as the dots' direction and strength of motion, colour, and the monkey's choice are encoded in the manifold (Mante et al., 2013). **e** Population activity in the mouse primary visual cortex in response to gratings of different orientations, indicated by color (Stringer et al., 2021). The panel is adapted from Jazayeri and Ostojic (2021)

statistical structures or properties they capture and preserve. It is important to select the type of method most suitable for the type of neural data being analyzed, as this may have significant impact on the resulting scientific interpretation.

In this section, we describe some of the more common manifold learning methods in use in the neuroscience literature. To make a fair and informative comparison of each algorithm, we implemented them and applied them each to a number of different neural datasets, as described in Section 3. In order to facilitate comparison of the assumptions made by the different algorithms, we have attempted to adopt (where possible) a common mathematical terminology throughout; this is summarised in Table 1.

2.1.1 Linear method

Linear manifold learning is accomplished by performing linear transformations of the observed variables that preserve certain optimality properties, which yield low-dimensional embeddings that are easy to interpret biologically. Some of the most common linear manifold learning techniques are discussed below.

2.2 Principal component analysis (PCA)

One of the most common linear manifold learning methods is Principal Component analysis (PCA) (Jolliffe, 2002; Jackson,

2005; Ivosev et al., 2008). It reduces the dimensionality of large datasets while preserving as much variability (or statistical information) as possible. To obtain a neural manifold **Y**, PCA performs an eigendecomposition on the covariance matrix of the neural activity matrix **X** in order to find uncorrelated latent variables that are constructed as linear combinations of the contributions of individual neurons, while successively maximising the variance. The computed eigenvectors,

Table 1 Summary of mathematical notation used throughout this manuscript

Notation	Description
N	Number of neurons
T	Number of time samples
t	Sample timestamp
$\mathbf{X}$	Population activity matrix
$\mathbf{x_t}$	Network activity state at timestamp t
k	Number of dimensions in the embedding
$\mathbf{Y}$	Manifold embedding matrix
$\mathcal{Y}$	Manifold - a topological structure within $\mathbf{Y}$
$\mathbf{D}$	Dissimilarity matrix (MDS)
$\mathbf{G}$	Graph of population activity
$\mathbf{\Lambda}$	Graph Laplacian (LEM)
$\mathbf{\Delta}$	Diagonal matrix (LEM)
$\mathbf{\Omega}$	Adjacency matrix (LEM)

or principal components (PCs), represent the directions of the axes that capture the most variance, and the corresponding eigenvalues yield the amount of variance carried by each PC.

Although PCA has been used to great effect to reduce the dimensionality of high-dimensional neural data (Churchland et al., 2010; Mazor & Laurent, 2005; Gao & Ganguli, 2015; Ahrens et al., 2012), its caveat is that it captures all types of variance in the data, including spiking variability, which may obscure the interpretation of latent variables. For example, neurons with higher firing rates typically exhibit higher variance (spike count variance being proportional to mean spike count for cortical neurons Tolhurst et al. (1983)), and therefore may skew the orthogonal directions found by PCA by accounting mostly for highly active neurons. Additionally, cortical neurons respond with variable strength to repeated presentations of identical stimuli (Tolhurst et al., 1983; Shadlen & Newsome, 1998; Cohen & Kohn, 2011). This variability is often shared among neurons, and such correlations in trial-to-trial responses can have a substantial effect on the amount of information encoded by a neuronal population. To minimise the effects of spiking variability, trial averaging or temporal smoothing of spike counts is usually done prior to performing PCA. However, this may not always be applicable to all analyses.

2.3 Multidimensional scaling (MDS)

Classical multidimensional scaling is another linear dimensionality reduction technique that aims to find a low-dimensional map of a number of objects (e.g. neural states) while preserving as much as possible pairwise distances between them in the high-dimensional space (Kruskal & Wish, 1978; Venna & Kaski, 2006; Yin, 2008; France & Carroll, 2010).

In the case of neural data, given an $N \times T$ activity matrix, $\mathbf{X}$, a $T \times T$ distance matrix, $\mathbf{D}$, is first obtained by measuring the distance between the population activity vectors for all t using a dissimilarity metric (Krauss et al., 2018). Examples of such metrics include both the Euclidean distance and the cosine dissimilarity. We will employ the latter for the MDS embeddings shown throughout this paper. For a pair of vectors $\mathbf{a},\mathbf{b}$ separated in angle by θ, the cosine dissimilarity is

$$D_{\mathbf{a},\mathbf{b}} = \cos(\theta) = 1 - \frac{\mathbf{a}.\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}.$$

The dissimilarity matrix $\mathbf{D}$ is formed from the cosine dissimilarities between the neural activity patterns $\mathbf{c}_t$ at each time t for all pairs of times measured. A lower-dimensional mapping, $\mathbf{Y} \in \mathfrak{R}^{k \times T}$, where $k << N$, is then found by minimising a loss function, called the *strain*, so that the mapped inter-point distances are as close as possible to the original distances in $\mathbf{D}$. From the eigen-decomposition of the distance matrix $\mathbf{D}$, the MDS components (i.e. dimensions) are revealed by the

eigenvectors, which are a linear combination of the distances across the T population activity vectors, while their respective eigenvalues report the amount of variance explained.

Multidimensional scaling has been used to study changes in neural population dynamics in both large-scale neural simulations (Phoka et al., 2012) and in neural population recordings (Luczak et al., 2009). It has been applied for characterizing glomerular activity across the olfactory bulb in predicting odorant quality perceptions (Youngentob et al., 2006), integrative cerebral cortical mechanisms during viewing (Tzagarakis et al., 2009), neuroplasticity in the processing of pitch dimensions (Chandrasekaran et al., 2007), emotional responses to music in patients with temporal lobe lesions (Dellacherie et al., 2011), and structural brain abnormalities associated with autism spectrum disorder (Irimia et al., 2018).

2.3.1 Non-linear methods

The algorithms described above can only extract linear (or approximately so) structure. Non-linear techniques, on the other hand, aim to uncover the broader non-linear structure of the neural population activity matrix $\mathbf{X}$. These insights, can come at the expense of weaker biological interpretation, as the discovered manifold is not given by a linear combination of some observed variable (e.g. individual neurons). Non-linear algorithms often attempt to approximate the true topology of the manifold $\mathcal{Y}$ within the reduced dimensionality representation $\mathbf{Y}$ by finding population-wide variables that discard the relationship between faraway points (or neural states) on $\mathbf{X}$ and focus instead on conserving the distances between neighbors.

2.4 Isomap

One of the most commonly used non-linear manifold learning algorithms is Isomap (Tenenbaum et al., 2000). Non-linear techniques aim to uncover the broader non-linear structure of the neural manifold embedding of the neural population activity, $\mathbf{X}$, by approximating the true topology of the neural manifold, $\mathcal{Y}$. To do so, Isomap first embeds $\mathbf{X}$ into a weighted graph $\mathbf{G}$, whose nodes represent the activity of the neuronal population at a given time $\mathbf{x}_t$, and the edges between them represent links to network states $\mathbf{x}_i, \mathbf{x}_j, \ldots$ that are the most similar to $\mathbf{x}_t$, i.e., its neighbors.

The k-nearest neighbors algorithm is usually used to estimate the neighbors for all network states $\mathbf{x}_1, \ldots, \mathbf{x}_T$. These neighboring relations are then encoded in a weighted graph with edges $d_{\mathbf{G}}(i,j)$ between neighboring states i, j that depends on the distance metric $d_{\mathbf{X}}$ used. $\mathbf{G}$ can then be used to approximate the true geodesic distances on the manifold $d_{\mathcal{Y}}(i,j)$ between any two points i, j (i.e. network states $\mathbf{x}_i, \mathbf{x}_j$)

by measuring their shortest path length $d_\mathbf{G}(i,j)$ in the weighted graph $\mathbf{G}$ using an algorithm such as Dijkstra's (Dijkstra et al., 1959).

MDS is then applied to the matrix of shortest path lengths within the graph $\mathbf{D_G} = \{d_\mathbf{G}(i,j)\}$ to yield an embedding of the data in a k-dimensional Euclidean space $\mathbf{Y}$ that best preserves the manifold's estimated geometry. The quality of the reconstructed manifold $\mathcal{Y}$ depends greatly on the size of the neighborhood search and the distance metric used to build $\mathbf{G}$. Isomap is a conservative approach that seems well suited to tackle the non-linearity inherent to neural dynamics and it has in fact been used in a variety of studies, even just for visualization purposes (Mimica et al., 2018; Chaudhuri et al., 2019; Sun et al., 2019).

2.5 Locally linear embedding (LLE)

LLE is another non-linear manifold learning technique that attempts to preserve the local geometry of the high-dimensional data $\mathbf{X}$, by separately analyzing the neighborhood of each network state $\mathbf{x}_t$, assuming it to be locally linear even if the global manifold structure is non-linear (Roweis and Saul, 2000). The local neighborhood of each network state is estimated using the steps described in Section 2.4 and is connected together to form a weighted graph $\mathbf{G}$. The location of any node i in the graph corresponds to the network state $\mathbf{x}_i$, which can then be described as a linear combination of the location of its neighboring nodes $\mathbf{x}_j, \mathbf{x}_k, ...$ These contributions are summarised by a set of weights, $\mathbf{W}$, which are optimised to minimise the reconstruction error between the high-dimensional network state $\mathbf{x}_i$ and the neighborhood linear estimation. The weight w_{ij} summarizes the contribution of the jth network state to the reconstruction of the ith state. To map the high-dimensional dataset $\mathbf{X}$ onto a lower-dimensional one $\mathbf{Y}$, the same local geometry characterized by $\mathbf{W}$ is employed to represent the low-dimensional network state $\mathbf{y}_i$ as a function of its neighbors $\mathbf{y}_j, \mathbf{y}_k, ...$

A significant extension of LLE was introduced that makes use of multiple linearly independent weight vectors for each neighborhood. This leads to a Modified LLE (MLLE) algorithm that is much more stable than the original (Zhang & Wang, 2007). Unlike Isomap, LLE preserves only the local geometry of the high-dimensional data $\mathbf{X}$, represented by the neighborhood relationship, so that short high-dimensional distances are mapped to short distances on the low-dimensional projection. In contrast, Isomap aims to preserve the geometry of the data at all scales, long distances included, possibly introducing distortions if the topology of the manifold is not estimated well. For neural data, though, Isomap has generally been preferred for its theoretical underpinnings and more intuitive approach.

2.6 Laplacian eigenmaps (LEM)

LEM, also referred to as Spectral Embedding, is another non-linear technique similar to LLE (Belkin, 2003). The algorithm is geometrically motivated as it exploits the properties of the neighboring graph $\mathbf{G}$ generated from the high-dimensional data $\mathbf{X}$, as in Section 2.4, to obtain a lower-dimensional embedding $\mathbf{Y}$ that optimally preserves the neighborhood information of $\mathbf{X}$.

In the graph $\mathbf{G}$, any two connected nodes (network states) i and j are connected by a binary edge using a neighborhood method as in 2.4, or by a weighted edge computed via a kernel parametrising the exponential relationship of the weights with respect to the distance between the nodes in the high-dimensional space $\mathbf{x}_i - \mathbf{x}_j$. The resulting graph edges form the adjacency matrix $\mathbf{\Omega}$ that is used, together with the diagonal matrix $\mathbf{\Delta}$ containing the degree of each node of $\mathbf{G}$, to obtain the graph Laplacian $\mathbf{\Lambda} = \mathbf{\Delta} - \mathbf{\Omega}$. The spectral decomposition of $\mathbf{\Lambda}$ reveals the structure and clusters on $\mathbf{G}$. The k eigenvectors with the smallest non-zero eigenvalues of $\mathbf{\Lambda}$ are, in fact, the k dimensions of the manifold embedding $\mathbf{Y}$. Similar to LLE, Laplacian eigenmaps preserve only the local geometry of the neural population activity $\mathbf{X}$ and are therefore more robust to the construction of $\mathbf{G}$. Indeed LEM has been successfully used to unveil behaviorally relevant neural dynamics (Rubin et al., 2019; Sun et al., 2019).

2.7 t-distributed stochastic neighbor embedding (t-SNE)

t-SNE is another non-linear method that aims to match local distances in the high-dimensional space $\mathbf{X}$ to the low-dimensional embedding $\mathbf{Y}$. This is obtained by first constructing a probability distribution over pairs of high-dimensional points $\mathbf{x}_i, \mathbf{x}_j$ in such a way that nearby points are assigned a higher probability while dissimilar points are assigned a lower probability. Then t-SNE defines a similar probability distribution over the points $\mathbf{y}_i, \mathbf{y}_j$ in the low-dimensional space, and it minimizes the Kullback-Leibler divergence between the two distributions (Van der Maatea & Hinton, 2008). The Euclidean distance is used in the original algorithm to evaluate the similarity between data points, but any appropriate metric can be employed as well. This method has been applied in a wide range of domains, from genomics to signal processing, including multiple neuroscientific settings (Dimitriadis et al., 2018; Panta et al., 2016). It usually considers up to three embedding dimensions for visualization constraints, and for exploiting the Barnes-Hut approximation, which reduces the computational cost to $O(N \log N)$ from $O(N^2)$. (Van Der Maaten, 2014).

2.8 Uniform manifold approximation and projection (UMAP)

UMAP is a non-linear manifold learning technique that is constructed from a theoretical framework based on topological data analysis, Riemannian geometry and algebraic topology (McInnes et al., 2018) and it has also been used on neural data both as an NML method (Tombaz et al., 2020) and for broader dimensionality reduction purposes Lee et al. (2021). It builds upon the mathematical foundations of LEM, Isomap and other non-linear manifold learning techniques in that it uses a *k* nearest neighbors weighted graphs representation of the data. By using manifold approximation and patching together local fuzzy simplicial set representations, a topological representation of the high-dimensional data is constructed. This layout of the data is then optimised in a low-dimensional space to minimize the cross-entropy between the two topological representations. Compared to t-SNE, UMAP is competitive in terms of visualization quality and arguably preserves more of the global dataset structure with superior run time performance. Furthermore, the algorithm is able to scale to significantly larger data set sizes than are feasible for t-SNE.

UMAP and t-SNE have recently being employed for visualizing high-dimensional genomics data and some distortion issues have been raised (Chari et al., 2021). Although this problem is particularly apparent with t-SNE, which tends to completely disregard the global structure of the data to find clusters, any reduction to too few dimensions with respect to the original high-dimensional space will inherently distort the topology of some of the data, as indicated by the Johnson-Lindenstrauss Lemma (Johnson and Lindenstrauss, 1984). These criticisms have been made with particular reference to genomics datasets, which are intrinsically higher-dimensional than the neural datasets which manifold learning has been applied to.

2.9 Probabilistic latent variable models

Another type of NML algorithm uses probabilistic latent variable models that construct a generative model for the data in terms of mapping a low-dimensional manifold or latent space to neural responses. This type of algorithm utilizes a probabilistic framework that performs temporal smoothing and dimensionality reduction simultaneously, allowing joint optimisation of the degree of smoothing and the relationship between the original high-dimensional data and the resulting low-dimensional neural trajectory (Yu et al., 2009). A good example is Gaussian Process Factory analysis (GPFA) that uses Gaussian processes and an additional explicit noise model to account for the different independent noise variances of different neurons (i.e., spiking variability). It is a set of factor analyzers that are linked together in the low-dimensional state space by a Gaussian process prior (Rasmussen & Williams, 2006), which allows for the specification of a correlation structure across the low-dimensional states at different time points. In cases where the neural time courses are believed to be similar across different trials, smooth firing rate profiles may be obtained by averaging across a small number of trials (Mazor & Laurent, 2005; Stopfer et al., 2003; Brown et al., 2005; Broome et al., 2006; Levi et al., 2005; Nicolelis et al., 1995), or by applying more advanced statistical methods for estimating firing rate profiles from single spike trains (DiMatteo et al., 2001; Cunningham et al., 2007) Similarly, Manifold Inference from Neural Dynamics (MIND) is a recently developed NML algorithm that aims to characterize the neural population activity as a trajectory on a non-linear manifold, defined by possible network states and temporal dynamics between them (Low et al., 2018; Nieh et al., 2021).

2.10 NML algorithms for trial-structured datasets

Importantly, model selection from a computed manifold can be greatly affected by the signal-to-noise ratio (SNR) of the initial input neural data. In many cases this has been overcome by using the square root transformation of spiking data and convolving it with a Gaussian filter to yield a smoothed instantaneous firing rate (Yu et al., 2009). In addition, multiple dimensionality reduction steps have also been used to enable more interpretable visualizations (LEM on LEM (Rubin et al., 2019), UMAP on PCA (Gardner et al., 2022)). Furthermore, the NML algorithms described and visualized up until now have been general use case algorithms, used to infer neural correlates from the data. However, in experiments where specific behaviors or decisions are time-locked and run across multiple trials, some of the NMLs described above have been augmented and optimised. These include, demixed Principal Component analysis (dPCA) (Kobak et al., 2016), Tensor Component analysis (TCA) (Cohen & Maunsell, 2010; Niell & Stryker, 2010; Peters et al., 2014; Driscoll et al., 2017), Cross-Validated PCA (cvPCA) (Stringer et al., 2019) and model-based Targeted Dimensionality Reduction (mTDR) (Aoi & Pillow, 2018). These NML algorithms exploit the trial nature of an experiment to discriminate signal from trial-to-trial variability or noise, enabling the experimenter to identify the principal components that maximally correspond to a stimulus or action.

2.11 ANN-based NML algorithms

Artificial neural networks (ANN) can also be employed for manifold learning as they have the potential to extract complex non-linear structure in high-dimensional data. Autoencoders exemplify this approach as they are designed to find an optimal encoding between a high-dimensional input

and a low-dimensional representation stored in their "bottle-neck" code layer, which preserves the information necessary to then reconstruct the original input from it. In this sense, auto-encoders can be thought of as a non-linear extension to PCA, where each node in the code layer is comparable to a PC. Delving deeper, the field of deep generative models such as variational auto-encoders (VAE) promise great potential at extracting low-dimensional structure in varied high-dimensional data, by constructing a stochastic model of the low-dimensional dynamics underlying the neural activity. Such methods have shown great results when inferring the neural activity trial-by-trial fluctuations, but some have raised the issue that the low-dimensional structure extracted from these models are often highly entangled and therefore, difficult to interpret (Pandarinath et al., 2018). To address these issues, VAEs that make use of external labels, such as behavioral variables or the passage of time, have been designed (Zhou & Wei, 2020). Lastly, addressing some of the shortcomings of VAEs such as interpretability, identifiability and generalizability, Consistent EmBeddings of high-dimensional Recordings using Auxiliary variables, or CEBRA, was developed. CEBRA uses an innovative approach that employs contrastive learning instead of a generative model to extract embeddings, this enables it to cope with strong data distribution shifts to yield consistent embeddings between experimental sessions, subjects and recording modalities (Schneider et al., 2022).

3 Manifold learning for the analysis of large-scale neural datasets

To demonstrate how neural manifold learning can be used in the analysis of large-scale neural datasets, we apply a number of linear and non-linear NML algorithms (PCA, MDS, LEM, LLE, Isomap, t-SNE and UMAP) to several datasets. The datasets were chosen to cover a number of different brain regions and a range of behavioral complexity. They consist of (i) two-photon calcium imaging of hippocampal subfield CA1 in a mouse running along a circular track (Section 3.1), taken from Go et al. (2021; ii) multi-electrode array extracellular electrophysiological recordings from the motor cortex of a macaque performing a radial arm goal-directed reach task (Section 3.2) from Yu et al. (2007); and (iii) single-photon "mini-scope" calcium imaging data recorded from the prefrontal cortex of a mouse under conditions where multiple task-relevant behavioral variables were monitored (Section 3.3), from Rubin et al. (2019). Lastly, we illustrate how manifold learning can be employed to characterize brain dynamics in a disease state such as Alzheimer's disease by applying these techniques to data simulated to reproduce basic aspects of dataset (i), augmented to incorporate pathology (Section 3.4).

3.1 Decoding from neural manifolds

To compare NML algorithms we evaluated the resulting manifolds according to behavioral decoding (reconstruction or classification) performance, ability to encode the high-dimensional activity (i.e reconstruction score) and intrinsic dimensionality. These quantifications make up a minor subset of the many manifold parameterization methods, of which we describe more Section 4.2). For hippocampal CA1 manifolds obtained by any of the NML methods, we computed decoding accuracy for the behavioral variable(s) as a function of the number of manifold embedding dimensions using an Optimal Linear Estimator (OLE) (Warland et al., 1997). This allows assessment of the number of dimensions necessary to encode the behavioral variable. We used a 10-fold cross-validation approach, i.e., training the decoder on 90% of the data and testing it on the remaining 10%. Decoding performance is calculated as the Pearson correlation coefficient between the actual and reconstructed behavioral variable, i.e. the mouse position, for the test data. To assess neural manifold information provided about animal behavior in the other two datasets, we built a logistic regression classifier (Hosmer et al., 1989); we evaluate its performance using the F1 score as a function of the number of manifold embedding dimensions used. The F1 score is defined as the weighted average of the precision (i.e., percentage of the results which are relevant) and recall (i.e., percentage of the total relevant results correctly classified by the algorithm), and ranges between 0 (worst) and 1 (best performance) (Blair, 1979).

3.2 Reconstruction of neural activity from a low-dimensional embedding

Another way to evaluate the degree of fidelity of the manifold embedding is to attempt to reconstruct the high-dimensional neural data from the low-dimensional embedding. This tells us how much has been lost in the process of dimensionality reduction. To obtain such a reverse mapping, we employed the non-parametric regression method originally introduced for LLE (Low et al., 2018; Nieh et al., 2021). We then obtained the reconstruction similarity by computing the Pearson correlation coefficient between the reconstructed and the original neural activity. To perform an element-wise comparison, the $N \times T$ neural activity matrices were concatenated column-wise into a single vector and the correlation coefficient calculated. To obtain the neural activity reconstruction score, we employed a 10-fold cross-validation strategy. Using 90% of the data from each session to learn the reverse mapping, the reconstruction was then evaluated on the remaining 10% the data. The final score was then obtained by averaging across folds.

3.3 Intrinsic manifold dimensionality

Estimating the number of required dimensions of the underlying manifold is a crucial part of manifold learning (Altan et al., 2021; Cunningham and Yu, 2014), as it helps one to acquire a conceptual idea of how complex the neuronal activity inside the manifold is. The intrinsic manifold dimensionality accounts for the number of independent (latent) variables necessary to describe the neural activity without suffering significant information loss (Jazayeri & Ostojic, 2021). However, it is difficult to estimate the dimensionality of neural manifolds, especially in the realistic condition of a noisy, non-linear embedding. A recent review provides a detailed evaluation of several dimensionality estimation algorithms when applied to high-dimensional neural data (Altan et al., 2021).

We acknowledge that any measure of dimensionality is strongly influenced by the timescale of the neural activity and by the size of the population recorded (Humphries, 2020), in this review we use the intrinsic dimensionality measure to compare the topologies captured by the different NML methods. To infer the manifold intrinsic dimensionality we employ a method related to the correlation dimension of a dataset (Grassberger and Procaccia, 1983). After applying NML to the original high-dimensional neural activity, for any given point (i.e., neural state) in the low-dimensional space, the number of neighboring neural states within a surrounding sphere as a function of the sphere's radius was calculated. The slope of the number of neighboring data points within a given radius on a log-log scale equals the exponent k in the power law $N(r) = cr^k$, where N is the number of neighbors and r is the sphere radius. k then provides an estimate of the intrinsic dimensionality of the neural activity. We fit the power law in the range between ~10 to 5000 neighbors or neural states, aiming to capture the relevant temporal scale for each task and related manifold (Rubin et al., 2019; Nieh et al., 2021). Note, in fact, that we have selected a particular range for each dataset, also depending on the number of time samples T available (details in respective figure captions).

3.4 Hippocampal neural manifolds

The hippocampus is well known to be involved in memory formation and spatio-contextual representations (Scoville & Milner, 1957; O'Keefe & Conway, 1978; Morris, 2006). NML has been recently applied to hippocampal neural activity by several authors, suggesting that the rodent hippocampal activity encodes various contextual, task-relevant variables, displaying more complex information processing than spatial tuning alone (Rubin et al., 2019; Nieh et al., 2021). Here, we re-analyze published data from a dataset comprising two-photon calcium imaging of hippocampal CA1 place

cells, to which previously only MDS had been applied Go et al. (2021), in order to compare manifolds extracted by different algorithms. We examine how different NML methods characterize the dynamics of hippocampal CA1 neurons along trajectories in low-dimensional manifolds as they coordinate during the retrieval of spatial memories.

The data was recorded from a head-fixed mouse navigating a circular track in a chamber floating on an air-pressurized table under a two-photon microscope (Fig. 3a). The mouse position (Fig. 3b) was simultaneously tracked using a magnetic tracker. The activity of 30 of 225 hippocampal CA1 cells recorded in the shown session is depicted in Fig. 3c. Of the 225 cells, 92 were classified as place cells by Go et al. (2021), and their normalized activity rate map, sorted according to place preference, is shown in Fig. 3d. Employing the activity of all 225 cells (both place and non-place selective), both linear (Fig. 3e) and non-linear (Fig. 3f) NML methods, revealed a cyclic pattern of transitions between network states corresponding to different locations along the circular track. The manifold formed by the dynamics of neural activity as the mice explored the full track forms a complete representation of the 2D structure of the track. We compared the algorithms in terms of decoding performance (Fig. 3g), neural activity reconstruction score (Fig. 3h) and intrinsic manifold dimensionality (Fig. 3i). All algorithms performed similarly in terms of the metrics considered, yielding almost the best possible decoding performance with just one manifold dimension and the best possible reconstruction similarity with two dimensions. Moreover, all manifolds were found to have a similar intrinsic dimensionality of around 2.

In this example, the behavioral complexity is approximately one dimensional (i.e., the mouse running in a single direction along a circular track can be mapped onto the single circular variable θ) and all NML methods produce embeddings which allow high decoding performance, with each algorithm already reaching near-maximum performance after incorporating only the first manifold dimension. This suggests that if the behavioral complexity is low and its information is broadly encoded within the neural population, any NML algorithm will yield broadly similar results. However, as we will see, this does not necessarily hold when complexity of the behavioral task is increased. In terms of the ability to capture the neural activity variance, the neural activity reconstruction score suggests that the highly non-linear tSNE and UMAP algorithms yield more informative low-dimensional embeddings.

3.5 Motor cortical neural manifolds

NML and dimensionality reduction techniques have also been applied to neural activity within motor and pre-motor cortical areas, in particular to suggest that the high variability observed in the single-neuron responses is disregarded when
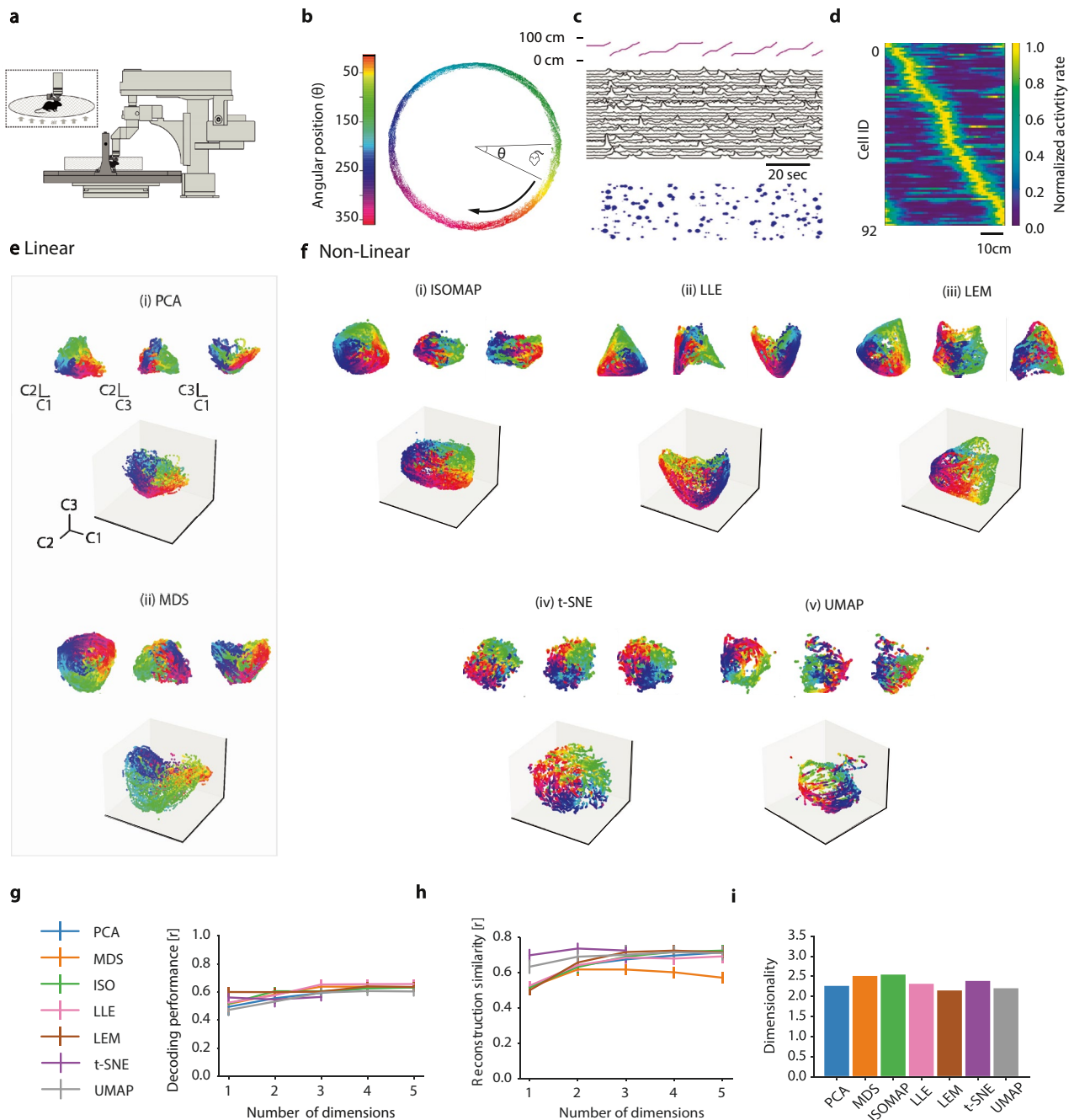
Fig. 3 Mouse hippocampalCA1 manifolds during spatial memory retrieval. Unless otherwise stated, panels adapted with permission from Go et al. (2021). **a** Schematic of experimental setup: head-fixed mouse navigates a floating circular track under a two-photon microscope. Inset: Close-up view of head-fixed mouse on floating track. **b** Spatial trajectories of the mouse, with θ (color-coded) denoting position on the track. **c** Top: Position along the track (cm), Middle: Ca^{2+} transients for 30 (of 225) randomly selected cells, Bottom: Rastergram showing events detected from the above calcium transients. Blue dots indicate time of Ca^{2+} transient onset, with dot area showing relative event amplitude (normalized per cell). **d** Normalized neural activity rate map for 92 place cells, sorted by spatial tuning around the track. **e - f** Linear vs non-linear manifold embeddings for all 225 cells (normalized). In each case the first three dimensions are visualized. Insets for each: projections on pairs of components C1 and C2 (upper left), C2 and C3 (upper middle), C1 and C3 (upper right). **e** Linear manifold embeddings (i) PCA, (ii) MDS, **f** non-linear manifold embeddings (i) Isomap, (ii) LLE, (iii) LEM, (iv) t-SNE, and (v) UMAP. (ii-v were not reproduced from Go et al. (2021). **g-i** Manifold evaluation metrics: (see: **g** Decoding performance (as used in Go et al. (2021)), **h** Neural activity reconstruction score, **i** Intrinsic manifold dimensionality

population dynamics are taken into account (Santhanam et al., 2009; Churchland & Shenoy, 2007; Churchland et al., 2012; Sussillo et al., 2015; Gallego et al., 2017). This approach has been the foundation of the "computation through dynamics" framework, which aims to characterize the neural population dynamics as trajectories along manifolds (Vyas et al., 2020). Rotational-like dynamics of motor cortex neural activity have been observed both in human and non-human primates (Churchland et al., 2012; Pandarinath et al., 2015), with stability over long periods of time (Gallego et al., 2020). Importantly, by furthering the understanding of neural population dynamics and variability, crucial steps can be made in improving performance of prosthetic devices that can be used to further enable those with nervous-system disease or injury in day-to-day tasks (Santhanam et al., 2009).

We applied NML to data from Yu et al. (2007) and Chestek et al. (2007) recorded in the caudal premotor cortex (PMd) and rostral primary motor cortex (M1) of rhesus macaques during a radial arm goal-directed reaching task (Fig. 4a). Neural data was collected for 800 successful repeats, with 100 trials for each of the 8 reaching directions (Fig. 4b). We used NML to analyze the neural activity during the 100 ms time window before the onset of the reaching movement and investigated its tuning with respect to the reaching direction (Santhanam et al., 2009). The manifold embeddings obtained using different NML methods, both linear (Fig. 4e) and non-linear (Fig. 4f), revealed different types of structures with data points clustering according to the monkey's target endpoint. All the NML algorithms tested revealed lower-dimensional structures that discriminate between each of the behavioral states along a single dimension. We used the pre-movement neural manifold to classify the behavior into one of eight different goal directions. The two linear NML algorithms yielded the most behaviorally informative embedding, requiring only two dimensions to achieve best performance (Fig. 4g). All algorithms performed equally in terms of neural activity reconstruction similarity, with only one dimension being necessary to reconstruct the original neural activity patterns (Fig. 4h). The intrinsic dimensionality of the two linear embeddings, on the other hand, was the highest (Fig. 4i). Non-linear NML algorithms extracted lower-dimensional embedding at the cost of encoding less behavioral information.

Increasingly, NML has been used to analyze neural circuits implicated in decision-making such as the prefrontal cortex (PFC) (Kingsbury et al., 2019; Mante et al., 2013) and the anterior cingulate cortex (ACC) (Rubin et al., 2019). With these multi-function brain regions, neurons are widely known for their mixed selectivity, often capturing information relating to both stimuli features and decisions (Kobak et al., 2016; Rigotti et al., 2013; Fusi et al., 2016). This moves away from the notion of highly-tuned single cells and gives rise to a dynamical, population-driven code (Pouget et al., 2000) ideally suited to NML methods. This was nicely demonstrated by Rubin et al. (2019), who visualized and quantified how NML (in their case, LEM) could be used to discriminate neural states arising from different brain regions (ACC and hippocampal area CA1) during identical tasks. The neural activity was recorded in freely-behaving mice exploring a linear track and performing various behaviors such as drinking, running, turning, and rearing (Figure 5a-c). Building on their findings, we employed various alternative NML methods, both linear (Fig. 5d) and non-linear (Fig. 5e), that revealed different neural data structures, exhibiting clustering according to the animals' behavior, with PCA and LLE producing the least clustered visualizations. Evaluation of NML behavior classification performance (Fig. 5f), reconstruction similarity (Fig. 5g) and manifold intrinsic dimensionality (Figure 5i) revealed that although there was high variability in performance, non-linear algorithms generally out-performed the linear ones. Notably, two non-linear NML algorithms t-SNE and UMAP performed the best in terms of behavioral classification and ability to reconstruct the high-dimensional neural activity from the manifold embedding, with an intrinsic dimensionality between 2 and 3, inferring that behavior is the predominant type of information encoded by the ACC network during this task.

3.6 Analysis of neural manifolds in neurological disorders

NML potentially provides a valuable tool for understanding the biology of different brain disorders. As an example, NML can be used to characterize changes in neural manifolds for spatial memory during the progression of Alzheimer's disease (AD). In AD, the hippocampus and connected cortical structures are among the first areas to show pathophysiology. Hippocampal-dependent cognitive processes such as episodic memory are particularly and prominently affected at the behavioral level. However, it is not yet understood how the pathological markers of AD, such as amyloid-beta plaques and neurofibrillary tangles, lead to specific disruptions in the network-level information processing functions underpinning memory function. While single-cell properties are obviously affected, it is believed that network properties relating to the coordination of information throughout brain circuits involved in memory may be particularly at risk (Palop & Mucke, 2016). Neural manifold learning analysis methods may thus play a useful or even crucial role in disentangling the effects of these network alterations.

The formation of extracellular amyloid plaques causes aberrant excitability in surrounding neurons (Busche et al., 2008): cells close to amyloid plaques tend to become hyperactive, whereas those further away from the plaques tend to
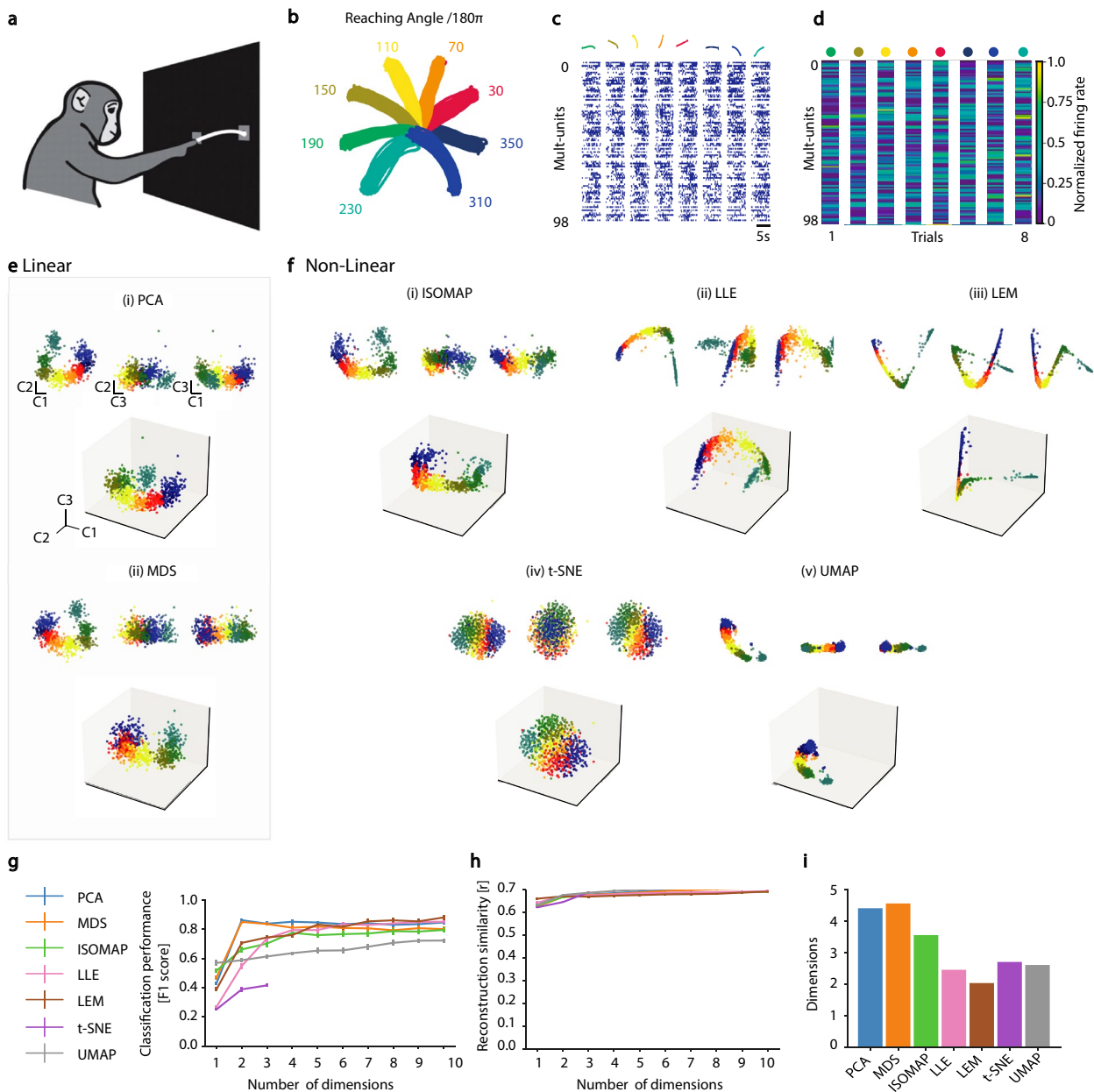
Fig. 4 Manifold analysis of motor cortical dynamicsduring a reaching task. **a** Monkey reaching task experimental setup. **b** Arm trajectories and target reaching angles - schematic and coloured trajectories adapted from Yu et al. (2007). **c-d** Neural data processing steps (adapted from Santhanam et al. (2009)) **c** Top: X,Y position of monkey arm over eight individual trials coloured according to (b). Bottom: Corresponding raster plot of spike trains recorded simultaneously from 98 neural units, sampled every 1 ms. Units include both well isolated single-neuron units (30% of all units), as well as multineuron units. **d** Population activity vector yielding the instantaneous firing rates of the same neural units in (c) for the 100 ms before the monkey reached out its arm (separated and coloured by trial type/arm direction as in c). **e - f** Linear vs non-linear manifold embeddings (normalized, only first three dimensions visualized). Insets: projections on pairs of components, C1 and C2 (upper left), C2 and C3 (upper middle), C1 and C3 (upper right). **(e)** Linear manifold embeddings (i) PCA, (ii) MDS. **f** non-linear manifold embeddings (i) Isomap, (ii) LLE, (iii) LEM, (iv) t-SNE, and (v) UMAP **g-i** Manifold evaluation metrics **g** Decoding performance [F1 score], **h** Neural activty reconstruction score [r], **i** Intrinsic manifold dimensionality

become hypoactive (or relatively silent). This aberrant neuronal excitability could potentially have a substantial effect on the neural manifold, and conversely, the neural manifold may provide a way to assess the overall impact of network

abnormalities. To demonstrate this, we simulated neural data for the (Go et al., 2021) circular track experiment described in Fig. 3) using parameters outlined by (Busche et al., 2008), i.e., the percentages of hyperactive (α) and hyperactive cells
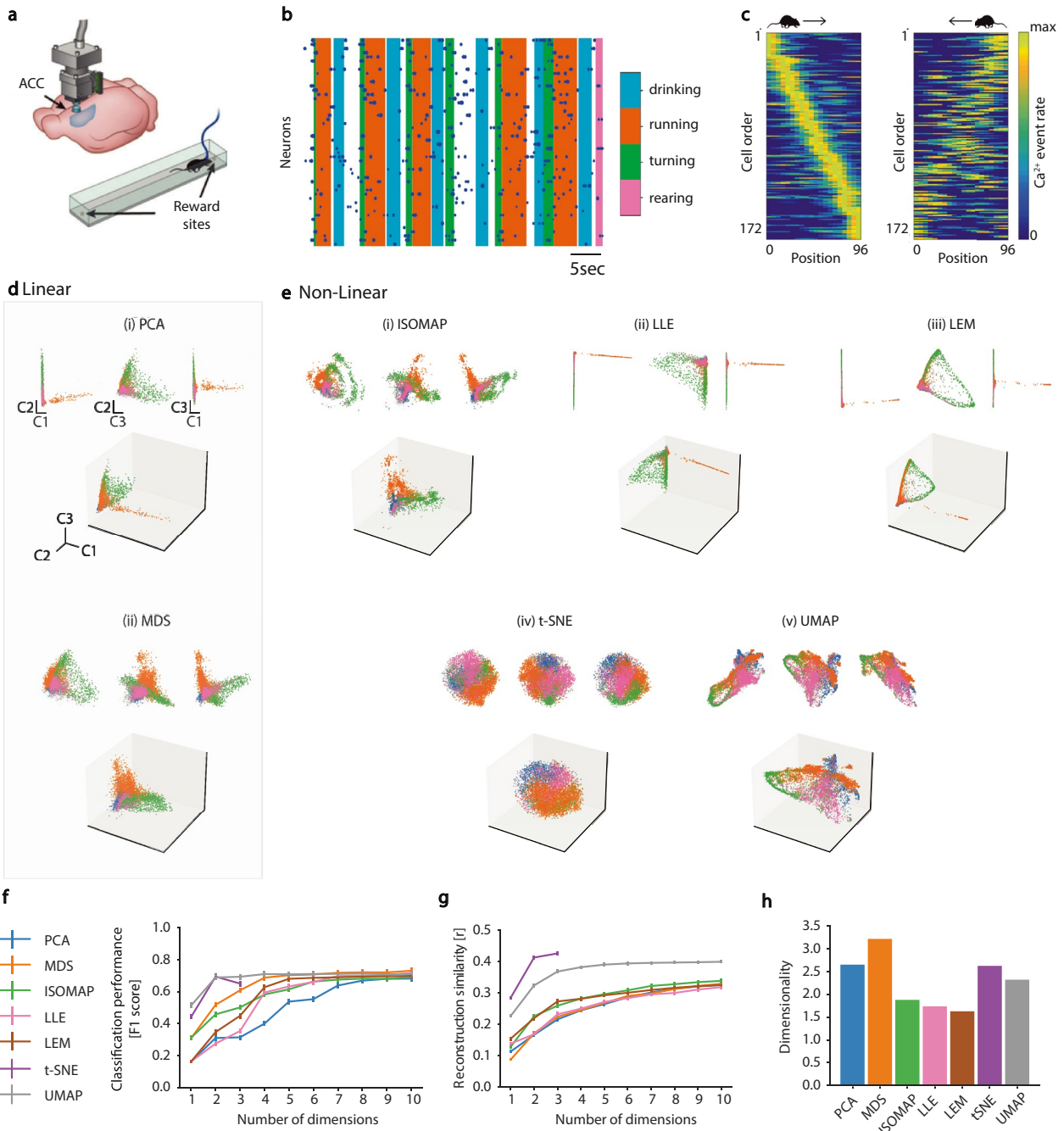
Fig. 5 Manifold analysis of prefrontalcortical dynamics during behavior. **a-c** Schematic of experimental setup from Rubin et al. (2019). The mouse navigates a linear track with rewards at both ends. **b** Raster plot of the neural activity during multiple behaviors (adapted from Rubin et al. (2019)): (1) drinking, (2) running, (3) turning, and (4) rearing. **c** Calcium event rate as the mouse runs to the right (left panel) and to the left (right panel) of the linear track (from Rubin et al. (2019)). **d-e** Linear vs non-linear manifold embeddings (normalized). In each case the first three dimensions are visualized: Insets for each: projections on pairs of components: C1 and C2 (upper left), C2 and C3 (upper middle), C1 and C3 (upper right). **d** Linear manifold embeddings (i) PCA, (ii) MDS, **e** non-linear manifold embeddings (i) Isomap, (ii) LLE, (iii) LEM, (iv) t-SNE, and (v) UMAP. **f-h** Manifold evaluation metrics **f** Decoding performance [r], **g** Reconstruction score [r], **h** Intrinsic manifold dimensionality

(β) in a given FOV, while conforming to previous studies that observed more than 20% of neurons becoming hyperactive (Busche et al., 2008; Busche & Konnerth, 2015)

in transgenic AD mice. This yielded a population of hippocampal CA1 cells (with 50% normal cells, 21% hyperactive cells and 29% hypoactive cells). Hypersynchrony, which

has also been observed in AD (Palop & Mucke, 2016), was not simulated for the purposes of this exercise, to maintain simplicity. This disease model mimics the neural activity of hippocampal CA1 cells of an AD mouse (e.g. a mouse over-expressing amyloid precursor protein, (Oakley et al., 2006)) running on a circular track; we compare it with a simulated control model representing a healthy wild-type litter-mate (Fig. 6a, b). Using MDS and UMAP as exemplars (Fig. 6c, d, respectively), NML shows how the aberrant excitability of neurons distorts the neural manifolds for spatial memory in AD (bottom panel) with respect to the healthy control model (top panel). While the topological structure remains relatively intact, the boundaries of the set become fuzzier, akin to adding noise to the system. To further evaluate this, we examined manifold parameterization measures including decoding performance, neural reconstruction similarity and manifold intrinsic dimensionality for both NML approaches (Fig. 6e-m). In both cases, the AD model's manifold embeddings display worse performance than the control model, although UMAP was able to recover a more behaviorally informative and lower-dimensional embedding than MDS.

To investigate the effects of the proportions of aberrant cells on the neural manifolds, we simulated the same models with varying percentages of normal, hyperactive (α) and hypoactive (β) cells. As expected, the hyperactivity of cells close to amyloid plaques (which tends to add noise) distorts the neural manifolds more than the hypoactive cells (which is more akin to having a decreased number of neurons in the manifold), resulting in less clustering of neural states in the manifold space. This is consistent with their effects on performance in terms of the given manifold measures (Fig. 6f-g, i-j, l-m).

We suggest that on this basis, NML techniques exhibit potential for helping to improve our understanding of network-level disruption in disease phenotypes.

4 Discussion

4.1 Comparison of neural manifold algorithms

With the increase in the numbers of simultaneously recorded neurons facilitated by emerging recording technologies, the demand for methods that enable reliable and comprehensively informative population-level analysis also increases. Multivariate analyses of large neuronal population activity thus require the use of more efficient computational models and algorithms that are geared towards finding the fewest possible population-level features that best describe the coordinated activity of such large neuronal populations. These features must be interpretable and open to parameterization that enables inherent neural mechanisms to be described (as discussed in the next section).

In this review, we give insights into how NML facilitates accurate quantitative characterizations of the dynamics of large neuronal populations in both healthy and disease states. However, assessing exactly how reliable and informative NML methods are remains a challenge and is open to interpretation. An ideal NML method must be able to adapt to different datasets that span functionally different brain regions, behavioral tasks, and number of neurons being analyzed, while considering different noise levels, timescales, etc.

In many studies, linear NML methods (such as PCA and classical MDS) have been preferred, and are widely used throughout the literature because of computational simplicity and clear interpretation. However, these methods have not been sufficient when the geometry of the neural representation is highly non-linear, which might for instance be induced by high behavioral task complexity (Jazayeri & Ostojic, 2021) (see Fig. 5). As shown in (Rubin et al., 2019), low SNR can also affect the quality of some linear NML embeddings. They compared PCA and LEM algorithms and revealed that LEM required fewer neurons and lower firing rates for accurate estimation of the internal structure. To address the challenges of noisy data, many solutions have been suggested, including the development of novel linear algorithms, such as dPCA (Kobak et al., 2016), cvPCA (Stringer et al., 2019), and mTDR (Aoi & Pillow, 2018). These methods enable greater discrimination between specific clusters of neural activity, but they do rely heavily on neural responses acquired during multiple trials of time-locked behavior. However, these may not always be available, and whilst it is possible to tweak certain algorithmic parameters or fit them to manipulated forms of the data (e.g. through time-warping), the end result is an increase in computational complexity and possibly harder to interpret results.

Conversely, non-linear methods, in general, trend towards better generalization across datasets. This is particularly true for UMAP and t-SNE (see Figs. 3, 4 and 5). To discriminate clusters of neural activity, UMAP and t-SNE have been shown to be the most powerful non-linear NML techniques. However, they might not always yield an embedding that is representative of the true high-dimensional geometry. These issues are particularly relevant for inherently very high-dimensional data, such as genomics data (Chari et al., 2021). In fact, it is debated whether a low number of latent dimensions are enough to drive neural population activity and explain its variability across a range of circumstances such as brain region and task (Chaudhuri et al., 2019; Rubin et al., 2019; Nieh et al., 2021; Churchland et al., 2012; Gallego et al., 2020), or if higher dimensionality are required (Stringer et al., 2019; Humphries, 2020).

Finally, a key consideration when selecting any NML method is whether its function is for neural data
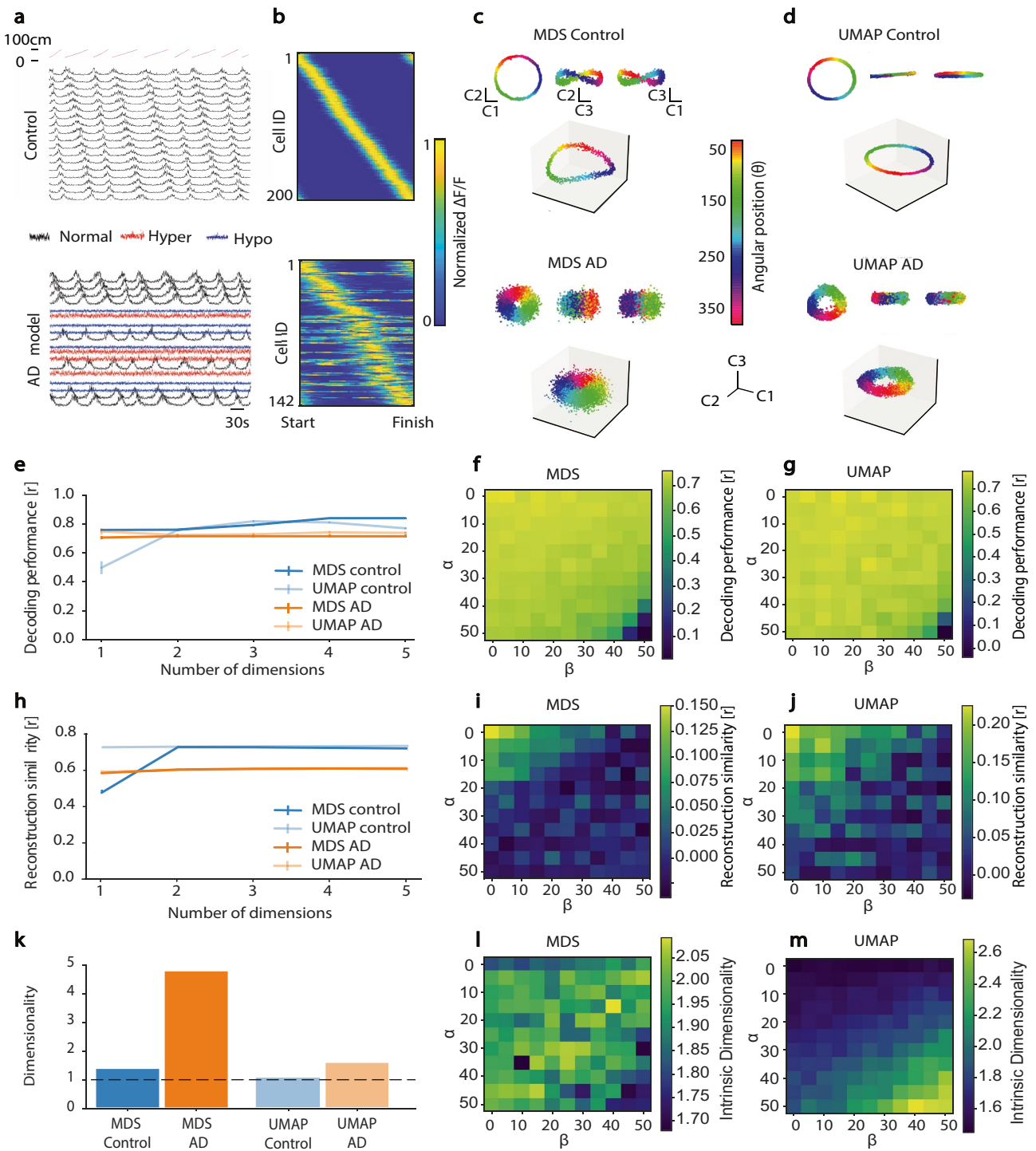
Fig. 6 Changes in neural manifolds during neurodegenerative disease states. **a** Example simulated calcium traces for a mouse traversing a circular track (simulation of real experiment in Fig. 3). Top: 20 of 200 "normal" CA1 place cells (magenta traces on top show mouse trajectory). Bottom: 20 of 200 AD-affected place cells that fall into three categories: Normal (black), Hyperactive (red) and Hypoactive (blue). **b** Normalized activity rate maps for 200 healthy and 142 (out of 200) AD hippocampal CA1 place and hyperactive cells (respec-tively) sorted by location of maximal activity. **c-d** 3D projection in the manifold space using (c) MDS and (d) UMAP for both healthy control and AD simulated models. Inset: 2D projections onto pairs of dimensions. **e-m** Evaluation metrics for MDS and UMAP, Control vs AD and shown as a function of varying percentages of normal, hyper-active (α) and hypoactive (β) cells. **e-g** Decoding performance [r]. **h-j** Reconstruction similarity [r]. **k-m** Intrinsic dimensionality

visualization or for structural discovery and/or quantification. For instance, many of the non-linear algorithms described above, such as UMAP, t-SNE have excellent capabilities in visualizing high-dimensional data and provide an interpretable global structure. Similarly, VAEs present themselves as great candidates for discriminating global clusters. However, obtaining a favorable global structure can often be at the expense of a loss to local structure (Chari et al., 2021) and therefore can distort any downstream structural quantification. To address these shortcomings one can use algorithms that either provide a linear solution such as PCA-based methods (Gardner et al., 2022) or that focus on local structure such as LEM (Rubin et al., 2019). ANNs may add further possibilities, with methods such as CEBRA tackling model identifiability and generalization (Schneider et al., 2022).

One message that can be taken away from our comparison of neural manifold approaches is that linear methods do provide substantial insight over a wide range of problem domains (often consistent with what is provided by non-linear methods), and even in areas where they fail to correctly capture the low-dimensional topological structure of the manifold, they can still provide insight verifiable through decoding of behavior.

4.2 Parameterization of neural manifolds

A key issue for NML approaches is that after the neural manifold has been determined, it normally needs to be analyzed further to answer a given scientific question. This process, known as parameterization, would ideally involve extraction of the equations of the manifold itself, however, out of feasibility may also be based on a more limited quantification of manifold properties. Depending on the type of neural data at hand and the scientific hypothesis to be tested, it's important to select the most suited and informative NML algorithm and parameterization measures. One challenge that has to be kept in mind is that the resulting manifolds, although often visualized in two or three dimensions for convenience, frequently involve many more dimensions (e.g. even a low-dimensional neural manifold from a recording from many hundreds of neurons might involve tens of dimensions), and many approaches that may work well in two or three dimensions do not necessarily scale straightforwardly to higher-dimensional representations.

Where transitions in the dynamics of neural states are observed over time, one approach that is useful is recurrence analysis (Marwan et al., 2007; Wallot, 2019). Recurrence analysis is a non-linear data analysis that is based on the study of phase space trajectories, where a point or element in phase space represents possibles states of the system. It is a powerful tool in characterizing the behavior of a dynamical system, especially when observing how states change over

time. A recurrence plot measures and visualizes the recurrences of a trajectory of a system in phase space. Note that although the recurrent plot is inherently two-dimensional, recurrence analysis can describe dynamics in an arbitrary dimensional space - a key feature of this technique. As the respective phase spaces of two systems change, recurrence plots allow quantification of their interaction and tell us when similar states of the underlying system occur. The time evolution of trajectories can be observed in the patterns depicted in the recurrence plots, and these patterns are associated with a specific behavior of the system, such as cyclicity and similarity of the evolution of states at different epochs. Different measures can be used to quantifying such recurrences based on the structure of the recurrence plot, including recurrence rate, determinism (predictibility), divergence, laminarity, and Shannon entropy (Webber & Marwan, 2015; Wallot, 2019). These measures might be especially useful when observing neuronal responses to stimuli over repeated trials under different conditions (such as healthy and diseased states).

Another way to compare neural manifolds under different conditions is to directly quantify the similarity of trajectories in the neural manifold space (Cleasby et al., 2019; Ding et al., 2008; Alt, 2009). The most commonly used measures of trajectory similarity are Fréchet distance (FD) (Fréchet, 1906; Besse et al., 2015), nearest neighbor distance (NND) (Clark & Evans, 1954; Freeman et al., 2011), longest common subsequence (LCSS) (Vlachos, 2002), edit distance for real sequences (EDR) (Chen & Ng, 2004; Chen et al., 2005) and dynamic time warping (DTW) (Toohey, 2015; Long & Nelson, 2013). Computing such measures may allow us to characterize the differences (or similarities) among neural manifolds obtained from different models. This type of analysis may be especially useful when comparing models in health and disease states across age groups, or when comparing neural manifold representations of the same model across different environments and conditions.

Another useful measure is manifold trajectory tangling (Russo et al., 2018). Tangling is a simple way of determining whether a particular trajectory could have been generated by a smooth dynamical flow field; under smooth dynamics, neural trajectories should not be tangled. Two neural states are thus tangled if they are nearby but associated with different trajectory directions. High tangling implies either that the system must rely on external commands rather than internal dynamics, or that the system is flirting with instability (Russo et al., 2018). Thus, tangling can be compared across all times and/or conditions. It is especially useful when characterizing neural dynamics over the course of learning or development. For example, it may be interesting to examine whether neural circuits adopt network trajectories that are increasingly less tangled when learning a new skill and with increasing performance. Conversely, it may also

be interesting to investigate whether pathological conditions may be associated with increased tangling during learning.

In cases where we characterize task-specific responses, e.g., during repetitive, trial-structured motor movements, determining whether these neural responses are consistent with the hypothesised computation is important. Hence, the consistency in the geometry of the population response and/or the time-evolving trajectory of activity in neural state space (i.e., manifold space) must be examined. For such task-specific trials, manifold trajectories are considered consistent if the trajectory segments trace the same path and do not diverge. One measure that can be used for such analysis is manifold trajectory divergence (Russo et al., 2020). In contrast to manifold trajectory tangling, which assesses whether trajectories are consistent with a locally smooth flow field, manifold trajectory divergence assesses whether similar paths eventually separate, smoothly or otherwise. A trajectory can have low tangling but high divergence, or vice versa. Multiple features can contribute to divergence, including ramping activity, cycle-specific responses, and different subspaces on different cycles. Thus, manifold trajectory divergence provides a useful summary of a computationally relevant property, regardless of the specifics of how it was achieved (Russo et al., 2020).

Lastly, topological data analysis (TDA), which comprises methods such as persistent cohomology, aim to rigorously characterize the topological structure of the data. These methods are quite sensitive to noise levels and might not be adequate to investigate functionally complex neural systems, but they succeed in revealing the topology of simpler systems such as the head direction, grid cell systems and linear track representation (Chaudhuri et al., 2019; Gardner et al., 2022; Rubin et al., 2019). Similarly, Spline Parameterization for Unsupervised Decoding (SPUD) is a powerful method to characterise the manifold dynamics, which stems from a rigorous analysis of the manifold topological structure (Chaudhuri et al., 2019).

An increasing body of work using large-scale neural recording technologies is pointing towards the viewpoint that neural manifolds during spontaneous activity may be relatively high dimensional (Avitan & Stringer, 2022) [reviewed in], in comparison to the low-dimensional picture of spontaneous activity that had emerged from earlier recording technologies such as single-unit extracellular electrophysiology and functional magnetic resonance imaging. This of course provokes the question: at what dimensionality do neural manifold approaches cease to become useful? Notably, the datasets we have used as illustrations in this survey paper are (reflecting the field) intrinsically relatively low-dimensional in nature, ranging from a mouse running around a constrained environment to a monkey moving its arm in a fixed set of trajectories. Free-ranging, real-world behaviour is obviously higher dimensional in nature. Do

manifold approaches cease to become useful in such cases? We would argue not, as long as the dimensionality of the manifold is much lower than the dimensionality of the neural recording (essentially determined by the number of cells that can be recorded, and which can exceed 10,000 with recently introduced technologies). Certainly two and three-dimensional visualisations may fail to capture interesting aspects of the data (which may in many cases be true already for 10-dimensional datasets). However, as we have pointed out in this section, there are many approaches for analysis of the resulting manifolds, such as recurrence plots, trajectory similarity measures, and manifold trajectory tangling, that work for higher dimensional manifolds. How these properties relate back to the predictions made by computational models of brain function is so far less clear. This area is ripe for further theoretical work.

4.3 Is neural manifold learning useful for understanding neurological disorders?

Neural manifold learning offers the prospect of aiding in our understanding of circuit neuropathologies. For instance, in mouse models of Alzheimer's disease, amyloid or tau pathologies result in changes in cortical circuitry, which are particularly evident in the hippocampus and connected structures. However, the effect of these pathologies on the dynamics of neural circuits involved in spatial and working memory at the network level are still not well understood. By analyzing the changes in the geometry of population responses and the time-evolving trajectory of activity of the associated hippocampal-cortical circuits, we can compare the neural manifolds recovered from groups of mice of different ages and/or with different health states. In particular, we can use recurrence analysis to determine whether neural states recur upon presentation of the same stimuli over repeated trials at different times across different ages of the models. Neural manifold similarity, tangling and divergence measures may also be computed to evaluate the differences (or similarity) of movement trajectories in the neural manifold space. Together, these measures can be used to compare the manifold dynamics across different age groups in both healthy and disease states.

In this review, we have demonstrated that neural manifold learning methods provide a powerful toolbox for understanding how populations of neurons coordinate in representing behaviorally relevant information. While the cellular and molecular pathologies underlying a variety of neurological disorders such as Alzheimer's Disease, Parkinson's Disease and Frontotemporal Dementia are at least beginning to be well understood, how they translate into network dysfunction and thus, into cognitive and behavioral deficits is not. Neural manifold learning techniques, in combination with new experimental technologies allowing us to record the

activity of many thousands of neurons simultaneously during behavioral tasks, potentially may allow us to assay the dynamics and Gestalt representations underlying cognition and behavior at the level of entire neural circuits. This could in turn lead to improved understanding of disease processes and more sensitive tests of therapeutic effect.

Code availability Code to generate some of the figures in this paper is available at: https://github.com/schultzlab/Neural_Manifolds

Declarations

Conflicts of interest The authors declare that there are no conflicts of interest.

References

Ahrens, M. B., Li, J. M., Orger, M. B., et al. (2012). Brain wide neuronal dynamics during motor adaptation in zebrafish. *Nature, 485*(7399), 471–477.

Alt, H. (2009). The computational geometry of comparing shapes. In: Efficient Algorithms. *Springer*, p 235–248

Altan, E., Solla, S. A., Miller, L. E. et al. (2021). Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *PLoS Computational biology 17*(11), e1008591

Aoi, M. C., & Pillow, J. W. (2018). Model-based targeted dimensionality reduction for neuronal population data. *Advances in Neural Information Processing Systems, 31*, 6690–6699.

Avitan, L., & Stringer, C. (2022). Not so spontaneous: Multi-dimensional representations of behaviors and context in sensory areas. *Neuron.* https://doi.org/10.1016/j.neuron.2022.06.019

Belkin, M. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation, 15*(6), 1373–1396.

Besse, P., Guillouet, B., Loubes, J. M., et al. (2015). Review and perspective for distance based trajectory clustering. arXiv:1508.04904

Blair, D. C. (1979). Information Retrieval, 2nd Edition. *Journal of the American Society for Information Science.*

Bouchard, K. E., Mesgarani, N., Johnson, K., et al. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(7441), 327–332.

Briggman, K. L., Abarbanel, H. D., & Kristan, W. B. (2005). Optical imaging of neuronal populations during decision-making. *Science, 307*(5711), 896–901.

Broome, B. M., Jayaraman, V., & Laurent, G. (2006). Encoding and decoding of overlapping odor sequences. *Neuron, 51*(4), 467–482.

Brown, S. L., Joseph, J., & Stopfer, M. (2005). Encoding a temporally structured stimulus with a temporally structured neural representation. *Nature neuroscience, 8*(11), 1568–1576.

Busche, M. A., & Konnerth, A. (2015). Neuronal hyperactivity-a key defect in alzheimer's disease? *Bioessays, 37*(6), 624–632.

Busche, M. A., Eichhoff, G., Adelsberger, H., et al. (2008). Clusters of hyperactive neurons near amyloid plaques in a mouse model of Alzheimer's disease. *Science, 321*(5896), 1686–1689.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative neurology and neuroscience, 25*(3–4), 195–210.

Chari, T., Banerjee, J., & Pachter, L. (2021). The specious art of single-cell genomics. bioRxiv:0825457696.

Chaudhuri, R., Gerçek, B., Pandey, B., et al. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience, 22*, 1512–1520.

Chen, L., & Ng, R. (2004). On the marriage of lp-norms and edit distance. *Proceedings of the Thirtieth international conference on Very large data bases-Volume, 30*, 792–803.

Chen, L., Özsu, M. T., & Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In: *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pp 491–502.

Chestek, C. A., Batista, A. P., Santhanam, G., et al. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *Journal of Neuroscience, 27*(40), 10742–10750.

Chung, S., & Abbott, L. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. *Current opinion in neurobiology, 70*, 137–144.

Churchland, M., Cunningham, J., Kaufman, M. T., et al. (2010). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron, 68*(3), 387–400.

Churchland, M., Cunningham, J., Kaufman, M., et al. (2012). Neural population dynamics during reaching. *Nature, 487*, 51–56.

Churchland, M. M., & Shenoy, K. V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *Journal of Neurophysiology, 97*(6), 4235–4257.

Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology, 35*(4), 445–453.

Cleasby, I. R., Wakefield, E. D., Morrissey, B. J., et al. (2019). Using time-series similarity measures to compare animal movement

trajectories in ecology. *Behavioral Ecology and Sociobiology, 73*(11), 1–19.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience, 14*(7), 811.

Cohen, M. R., & Maunsell, J. H. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. *Journal of Neuroscience, 30*(45), 241–253.

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience, 17*(11), 1500–1509.

Cunningham, J. P., Yu, B. M., Shenoy, K. V., et al. (2007). Inferring neural firing rates from spike trains using gaussian processes. *Advances in Neural Information Processing Systems, 20*, 329–336.

Dellacherie, D., Bigand, E., Molin, P., et al. (2011). Multidimensional scaling of emotional responses to music in patients with temporal lobe resection. *Cortex, 47*(9), 1107–1115.

Dijkstra, E. W., et al. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*(1), 269–271.

DiMatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika, 88*(4), 1055–1071.

Dimitriadis, G., Neto, J. P., & Kampff, A. R. (2018). t-sne visualization of large-scale neural recordings. *Neural Computation, 30*(7), 1750–1774.

Ding, H., Trajcevski, G., Scheuermann, P., et al. (2008). Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment, 1*(2), 1542–1552.

Driscoll, L. N., Pettit, N. L., Minderer, M., et al. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell, 170*(5), 986–999.

Elsayed, G. F., Lara, A. H., Kaufman, M. T., et al. (2016). Reorganization between preparatory and movement population responses in motor cortex. *Nature Communications, 7*(1), 1–15.

Feulner, B., & Clopath, C. (2021). Neural manifold under plasticity in a goal driven learning behaviour. *PLoS Computational Biology 17*(2), e1008621.

France, S. L., & Carroll, J. D. (2010). Two-way multidimensional scaling: A review. I*EEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and .Reviews) 41*(5), 644–661.

Fréchet, M. M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940), 22*(1), 1–72.

Freeman, R., Mann, R., Guilford, T., et al. (2011). Group decisions and individual differences: route fidelity predicts flight leadership in homing pigeons (columba livia). *Biology letters, 7*(1), 63–66.

Frost, N. A., Haggart, A., & Sohal, V. S. (2021). Dynamic patterns of correlated activity in the prefrontal cortex encode information about social behavior. *PLoS Biology, 19*(5), e3001235.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology, 37*, 66–74.

Gallego, J., Perich, M., Chowdhury, R., et al. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience, 23*, 1–11.

Gallego, J. A., Perich, M. G., Miller, L. E., et al. (2017). Neural manifolds for the control of movement. *Neuron, 94*(5), 978–984.

Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology, 32*, 148–155.

Gardner, R. J., Hermansen, E., Pachitariu, M., et al. (2022). Toroidal topology of population activity in grid cells. *Nature, 602*(7895), 123–128.

Go, M. A., Rogers, J., Gava, G. P., et al. (2021). Place cells in head-fixed mice navigating a floating real-world environment. *Frontiers in Cellular Neuroscience, 15*, 19.

Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters, 50*(5), 346.

Harvey, C. D., Coen, P., & Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature, 484*(7392), 62–68.

Hosmer, D. W., Jovanovic, B., & Lemeshow, S. (1989). Best subsets logistic regression. *Biometrics* pp. 265–1270.

Humphries, M. D. (2020). Strong and weak principles of neural dimension reduction. arXiv:2011.08088

Irimia, A., Lei, X., Torgerson, C. M. et al. (2018). Support vector machines, multidimensional scaling and magnetic resonance imaging reveal structural brain abnormalities associated with the interaction between autism spectrum disorder and sex. *Frontiers in Computational Neuroscience* p. 93.

Ivosev, G., Burton, L., & Bonner, R. (2008). Dimensionality reduction and visualization in principal component analysis. *Analytical chemistry, 80*(13), 4933–4944.

Jackson, J. E. (2005). *A user's guide to principal components*. John Wiley & Sons.

Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology, 70*, 113–120.

Johnson, W., & Lindenstrauss, J. (1984). Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics, 26*, 189–206.

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Kaufman, M. T., Churchland, M. M., Ryu, S. I., et al. (2014). Cortical activity in the null space: permitting preparation without movement. *Nature Neuroscience, 17*(3), 440–448.

Kingsbury, L., Huang, S., Wang, J., et al. (2019). Correlated neural activity and encoding of behavior across brains of socially interacting animals. *Cell, 178*(2), 429–446.

Kobak, D., Brendel, W., Constantinidis, C., et al. (2016). Demixed principal component analysis of neural population data. *eLife 5*, e10989.

Krauss, P., Metzner, C., Schilling, A., et al. (2018). A statistical method for analyzing and comparing spatiotemporal cortical activation patterns. *Scientific reports, 8*(1), 1–9.

Kruskal, J., & Wish, M. (1978). *Multidimensional Scaling*. Sage Publications.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nature Reviews Neuroscience, 3*(11), 884–895.

Lee, E. K., Balasubramanian, H., Tsolias, A., et al. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *eLife, 10*, e67490.

Levi, R., Varona, P., Arshavsky, Y. I., et al. (2005). The role of sensory network dynamics in generating a motor program. *Journal of Neuroscience, 25*(42), 9807–9815.

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science, 27*(2), 292–318.

Low, R. J., Lewallen, S., Aronov, D., et al. (2018). Probing variability in a cognitive map using manifold inference from neural dynamics. bioRxiv:418939.

Luczak, A., Barthó, P., & Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron, 62*(3), 413–425.

Machens, C. K., Romo, R., & Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal of Neuroscience, 30*(1), 350–360.

Mante, V., Sussillo, D., Shenoy, K. V., et al. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature, 503*(7474), 78–84.

Marwan, N., Romano, M. C., Thiel, M., et al. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5–6), 237–329.

Mazor, O., & Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron, 48*(4), 661–673.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv: 1802.03426

Meshulam, L., Gauthier, J. L., Brody, C. D., et al. (2017). Collective behavior of place and non-place neurons in the hippocampal network. *Neuron, 96*(5), 1178–1191.

Mimica, B., Dunn, B. A., Tombaz, T., et al. (2018). Efficient cortical coding of 3d posture in freely behaving rats. *Science, 362*(6414), 584–589.

Morris, R. (2006). Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience, 23*(11), 2829–2846.

Nicolelis, M. A., Baccala, L. A., Lin, R., et al. (1995). Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science, 268*(5215), 1353–1358.

Nieh, E. H., Schottdorf, M., & Freeman, N. W., et al. (2021). Geometry of abstract learned knowledge in the hippocampus. *Nature* pp. 1–5.

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron, 65*(4), 472–479.

Oakley, H., Cole, S. L., Logan, S., et al. (2006). Intraneuronal β-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial alzheimer's disease mutations: potential factors in amyloid plaque formation. *Journal of Neuroscience, 26*(40), 10129–10140.

O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research, 31*(4), 573–590.

Palop, J. J., & Mucke, L. (2016). Network abnormalities and interneuron dysfunction in alzheimer disease. *Nature Reviews Neuroscience, 17*(12), 777–792.

Pandarinath, C., Gilja, V., Blabe, C. H., et al. (2015). Neural population dynamics in human motor cortex during movements in people with als. *Elife, 4*(e07), 436.

Pandarinath, C., O'Shea, D. J., Collins, J., et al. (2018). Inferring single-trial neural population dynamics using sequential autoencoders. *Nature methods, 15*(10), 805–815.

Pang, R., Lansdell, B. J., & Fairhall, A. L. (2016). Dimensionality reduction in neuroscience. *Current Biology, 26*(14), R656–R660.

Panta, S. R., Wang, R., Fries, J., et al. (2016). A tool for interactive data visualization: application to over 10,000 brain imaging and phantom mri data sets. *Frontiers in Neuroinformatics, 10*, 9.

Peters, A. J., Chen, S. X., & Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity during motor learning. *Nature, 510*(7504), 263–267.

Phoka, E., Wildie, M., Schultz, S. R., et al. (2012). Sensory experience modifies spontaneous state dynamics in a large-scale barrel cortical model. *Journal of Computational Neuroscience, 33*(2), 323–339.

Pouget, A., Dayan, P., & Peter, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1*, 125–132.

Rajan, K., Harvey, C. D., & Tank, D. W. (2016). Recurrent network models of sequence generation and memory. *Neuron, 90*(1), 128–142.

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience, 17*(12), 1784–1792.

Rasmussen, C. E., & Williams, C. (2006). Gaussian processes for machine learning the mit press. Cambridge, MA.

Rigotti, M., Barak, O., Warden, M. R., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*(7451), 585–590.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*(5500), 2323–2326.

Rubin, A., Sheintuch, L., Brande-Eilat, N., et al. (2019). Revealing neural correlates of behavior without behavioral measurements. *Nature Communications, 10*(1), 1–14.

Russo, A. A., Bittner, S. R., Perkins, S. M., et al. (2018). Motor cortex embeds muscle-like commands in an untangled population response. *Neuron, 97*(4), 953–966.

Russo, A. A., Khajeh, R., Bittner, S. R., et al. (2020). Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation. *Neuron, 107*(4), 745–758.

Rust, N. C. (2014). Population-based representations. In: Gazzaniga MS, Mangun GR (eds) The Cognitive Neurosciences. MIT Press, chap 19, p 337.

Sanger, T. D., & Kalaska, J. F. (2014). Crouching tiger, hidden dimensions. *Nature Neuroscience, 17*(3), 338–340.

Santhanam, G., Yu, B., Gilja, V., et al. (2009). Factor-analysis methods for higher-performance neural prostheses. *Journal of Neurophysiology, 102*(2), 1315–1330.

Schneider, S., Lee, J. H., & Mathis, M. W. (2022). Learnable latent embeddings for joint behavioral and neural analysis. arXiv:2204.00673

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*(1), 11.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience, 18*(10), 3870–3896.

Stokes, M. G., Kusunoki, M., Sigala, N., et al. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron, 78*(2), 364–375.

Stopfer, M., Jayaraman, V., & Laurent, G. (2003). Intensity versus identity coding in an olfactory system. *Neuron, 39*(6), 991–1004.

Stringer, C., Pachitariu, M., Steinmetz, N., et al. (2019). High-dimensional geometry of population responses in visual cortex. *Nature, 571*(7765), 361–365.

Stringer, C., Michaelos, M., Tsyboulski, D., et al. (2021). High-precision coding in visual cortex. *Cell, 184*(10), 2767–2778.

Sun, G., Zhang, S., Zhang, Y., et al. (2019). Effective dimensionality reduction for visualizing neural dynamics by laplacian eigenmaps. *Neural Computation, 31*(7), 1356–1379.

Sussillo, D., Churchland, M. M., Kaufman, M. T., et al. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience, 18*(7), 1025–1033.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290*(5500), 2319–2323.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23*(8), 775–785.

Tombaz, T., Dunn, B. A., Hovde, K., et al. (2020). Action representation in the mouse parieto-frontal network. *Scientific reports, 10*(1), 1–14.

Toohey, K. (2015). Similaritymeasures: trajectory similarity measures. R package version 1.

Tzagarakis, C., Jerde, T. A., Lewis, S. M., et al. (2009). Cerebral cortical mechanisms of copying geometrical shapes: a multidimensional scaling analysis of fmri patterns of activation. *Experimental brain research, 194*(3), 369–380.

Van der Maaten, L., & Hinton. G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research, 9*(11).

Van Der Maaten, L. (2014). Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research, 15*, 3221–3245.

Venna, J., & Kaski, S. (2006). Local multidimensional scaling. *Neural Networks, 19*(6–7), 889–899.

Vlachos, M., Kollios, G., & Gunopulos, D. (2002). Discovering similar multidimensional trajectories. In: Proceedings 18th International Conference on Data Engineering, IEEE, pp 673–684.

Vyas, S., Golub, M. D., Sussillo, D., et al. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience, 43*, 249–275.

Wallot, S. (2019). Multidimensional cross-recurrence quantification analysis (mdcrqa)-a method for quantifying correlation between multivariate time-series. *Multivariate Behavioral Research, 54*(2), 173–191.

Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology, 78*(5), 2336–2350.

Webber, C., & Marwan, N. (2015). Recurrence quantification analysis. Theory and Best Practices.

Wu, A., Roy, N. A., Keeley, S., et al. (2017). Gaussian process based nonlinear latent structure discovery in multivariate spike train data. *Advances in Neural Information Processing Systems, 30*, 3496.

Yin, H. (2008). On multidimensional scaling and the embedding of self-organising maps. *Neural Networks, 21*(2–3), 160–169.

Youngentob, S. L., Johnson, B. A., Leon, M., et al. (2006). Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. *Behavioral neuroscience, 120*(6), 1337.

Yu, B. M., Kemere, C., Santhanam, G., et al. (2007). Mixture of trajectory models for neural decoding of goal-directed movements. *Journal of Neurophysiology, 97*(5), 3763–3780.

Yu, B. M., Cunningham, J. P., Santhanam, G., et al. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology, 102*(1), 614–635.

Zhang, Z., & Wang, J. (2007). MLLE: Modified locally linear embedding using multiple weights. *Advances in Neural Information Processing Systems, 19*, 1593–1600.

Zhou, D., & Wei, X. X. (2020). Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-vae. *Advances in Neural Information Processing Systems, 33*, 7234–7247.