

RESEARCH ARTICLE

# Exploring neural manifolds across a wide range of intrinsic dimensions

Jacopo Fadanni<sup>1</sup>, Rosalba Pacelli<sup>2</sup>, Alberto Zucchetta<sup>2</sup>, Pietro Rotondo<sup>3</sup>, Michele Allegra<sup>1,4\*</sup>

**1** Physics and Astronomy Department, University of Padova, Padova, Italy, **2** Istituto Nazionale di Fisica Nucleare, Sezione di Padova, Padova, Italy, **3** Department of Mathematical, Physical and Computer Sciences, University of Parma, Parma, Italy, **4** Padova Neuroscience Center, University of Padova, Padova, Italy

\* [michele.allegra@unipd.it](mailto:michele.allegra@unipd.it)



## Abstract

The rapid surge in the number of simultaneously recorded neurons demands reliable tools to explore the latent geometry of high-dimensional neural spaces. Within such spaces, neuronal activity typically lies on a subspace or manifold characterized by an intrinsic dimension (ID) that is much lower than the total number of recorded units. The ID can provide immediate information about the neural code, such as the minimum number of encoded variables and the relation between collective and individual neural activity. Existing studies rely on disparate and potentially unreliable ID estimators, which can contribute to conflicting reports of high-dimensional vs. low-dimensional manifolds. Here, we propose a robust and versatile pipeline for ID estimation, exploiting a local version of the full correlation integral estimator (IFCI). Being able to simultaneously cope with high dimensionality and non-linearity, IFCI overcomes some major limitations of common ID estimation methods. We prove the strength and accuracy of IFCI by applying it to synthetic benchmark data by Altan et al., 2019, where other methods typically underestimate the ID. We apply IFCI to study neural manifolds arising in recurrent neural networks trained on the 20 tasks of the well-known ‘cog-Task’ battery. Across tasks and training repetitions, IFCI uncovers a consistently low ID, which we show to be fundamentally related to the task structure. Finally, we apply IFCI to a reference experimental dataset by Stringer et al., 2019, comprising visual responses to a large set of natural images, strongly supporting previous reports that responses are organized in a high-dimensional manifold. IFCI has the potential to shed light on the current debate about the geometry of neural codes, and its dependence on structural constraints and computational goals in biological and artificial neural networks.

## OPEN ACCESS

**Citation:** Fadanni J, Pacelli R, Zucchetta A, Rotondo P, Allegra M (2026) Exploring neural manifolds across a wide range of intrinsic dimensions. *PLoS Comput Biol* 22(4): e1014162. <https://doi.org/10.1371/journal.pcbi.1014162>

**Editor:** Tatyana O Sharpee, Salk Institute, UNITED STATES OF AMERICA

**Received:** September 18, 2025

**Accepted:** March 23, 2026

**Published:** April 3, 2026

**Copyright:** © 2026 Fadanni et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** The source code and data used to produce the results and analyses presented in this manuscript are available on GitHub: <https://github.com/JFadanni/ExploringNeuralManifolds.git>.

**Funding:** This work received financial support from the Italian Ministry of University (<https://>

[www.mur.gov.it](http://www.mur.gov.it)), PRIN grant 2022HSLK9, "Unveiling the role of low dimensional activity manifolds in biological and artificial neural networks", CUP C53D23000740006/, CUP D53D23002220001 awarded to MA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

The dimensionality, or intrinsic dimension (ID), of neural manifolds is one of the key properties characterizing the collective organization of neural activity at a population level. However, rigorously estimating the ID is difficult: common methods tend to overestimate the ID in the presence of curvature, and underestimate it when large. Exploiting FCI, an ID estimator that is particularly effective in high-dimensional settings, we developed local FCI (IFCI), a pipeline for ID estimation that addresses these limitations. IFCI proves robust when tested against challenging benchmark data. We applied IFCI in two paradigmatic cases: data from artificial recurrent neural networks trained on simple tasks, and experimental data of neural responses to visual stimuli. In both cases, IFCI provides substantial support for widely debated hypotheses - low ID in the case of simple tasks and large ID for visual responses. IFCI could be used to characterize neural manifolds across a wide range of dimensions, offering a useful tool in the current debate about the dimensionality of neural activity.

## Introduction

Traditionally, brain circuits were analyzed by characterizing the functional role of individual units. The advent of large-scale neuronal recordings [1] opened the door to a different approach, focused on understanding the collective activity of all units in terms of representations and computation within abstract spaces [2,3]. Central to this perspective is the concept of *neural manifold* [4,5], the set of configurations visited by the system within the space defined by the activity of all units [6–10]. Visualizing, analyzing, and modeling neural manifolds requires identifying a minimum set of coordinates describing the manifold. The minimum number of coordinates is called the manifold's *intrinsic dimension (ID)*. The ID encapsulates the complexity of neuronal dynamics, in the classical sense of the relation between the individual units and the whole system [11]. The ID also bounds the possible number of cognitively and behaviorally relevant variables directly encoded by neurons at a population level.

Characterizing the ID of neural manifolds is essential to understand how collective neuronal activity supports the execution of cognitive tasks. Several studies investigating neural manifolds during simple instructed tasks reported a low ( $\lesssim 10$ ) ID [12–17], inspiring a large body of theoretical work aiming to explain how low-dimensional activity can emerge from structural constraints [18–21]. However, many other studies reported a high ID, typically in presence of complex stimulus spaces or loose task structure [22,23]. A thorough comprehension of what determines the geometry of neural manifolds, of which the ID is a key characteristic, is still missing. The specific type of computation or 'task' being performed is certainly a crucial factor, but not the only one: for any given task, neural networks may also need to strike an optimal trade-off between conflicting demands such as robustness (which favors low dimensional codes [24]) and ease of readout (which favors high-dimensional ones [25–30]).

A clearer understanding demands progress not only in neuronal recording techniques, but also in analysis methodology. In this context, a key gap is the lack of a fully reliable and shared methodology to measure the ID of neural manifolds [31].

Research on neural manifolds often employs linear ID estimation methods [14,15,20,22,23,26,27,32–40], which are based on the spectrum of the covariance matrix and use several criteria to select the number of ‘relevant’ or ‘significant’ dimensions. ID estimates by linear methods generally coincide with the dimension of the (minimal) linear subspace encompassing the manifold [41], rather than the manifold’s actual ID. As a result, they can significantly overestimate the ID in the case of non-linear, curved manifolds, despite the fact that the latter are quite common [42–44]. An alternative to linear methods is provided by geometric ID estimators [45–47], which are ultimately based on the principle that distances in the data follow scaling laws depending parametrically on the ID. The major shortcoming of these methods is that they are not robust to undersampling, requiring a sample size growing exponentially with the ID. As a consequence, they are generally unable to yield proper ID estimates when the ID is large ( $\gtrsim 10$ ). These shortcomings were highlighted in Ref [31], where simulated data were used to show that no common method can correctly identify the ID of non-linear, high-dimensional neural manifolds.

Here, we propose a pipeline to evaluate the ID of neural manifolds by leveraging the Full Correlation Integral (FCI) estimator [48]. The major advantage of FCI is its remarkable robustness with respect to undersampling, as it can estimate high IDs from samples whose size scales only linearly with the ID. On the other hand, FCI typically overestimates the ID in the presence of curvature, similarly to linear methods. To overcome this drawback, our pipeline employs a local version of FCI, called *local FCI (IFCI)*. We explore IFCI’s potential by applying it to artificial and real data sets: (i) neural manifolds generated by recurrent neural networks (RNN) trained on the ‘cog-task’ battery [49]. This task set, inspired by real tasks used in neurophysiological studies on non-human animals, involves basic cognitive processes such as working memory, inhibition, and context-dependent integration and it has become a sort of standard [50–54]. The simplicity of the stimulus-response patterns in RNNs is reflected in a strong and consistent expectation about the low dimensionality of the corresponding neural manifolds. (ii) a benchmark dataset of synthetic neural trajectories, embedded in high-dimensional space both linearly and non-linearly [31]. Trajectories are generated from an empirical distribution of firing rates, obtained from multi-electrode array recordings in the macaque primary motor cortex. As shown by Altan et al. [31], common linear and nonlinear estimators fail in the high-dimensional case, underestimating the dimension for  $ID \gtrsim 10$ . (iii) real data from calcium imaging of the primary visual cortex of mice [22]. This data set contains neural responses to both large (2,800 images) and small (32 images) stimulus spaces. The authors argued that responses are organized in a very high-dimensional manifold (a widely referenced result). Their conclusions, however, are entirely based on linear dimension estimators. For the cog-task battery, we find that our estimator works better than linear estimators and several geometric estimators, providing consistent estimates for different trainings of the networks. In the benchmark dataset, our estimator performs comparably with the best linear estimator (in the linear embedding case) and performs better than other methods (in the non-linear case). For the real dataset, our conclusions align with those of Stringer et al., supporting the existence of a high-dimensional manifold of visual responses.

This manuscript is organized as follows: in Methods, we provide a review of state-of-the-art ID estimators, including FCI, and illustrate the process to generate the synthetic datasets investigated. In the Results section, we present our pipeline (IFCI) and the results of its application to synthetic and real neural manifolds.

## Materials and methods

### Intrinsic dimension estimation

Consider a set  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_P\}$  of  $P$  points embedded in an ambient space of dimension  $N$ . Assuming that the points are contained, exactly or approximately, within a manifold or hypersurface of *intrinsic dimension*  $D < N$ , one can try to estimate  $D$ . There exists a wide array of ID estimation methods (for a classic review, see [55]). To be suitable for characterizing

neural activity manifolds, an estimator should ideally work well for non-linear, curved manifolds and a large range of dimensions, including high dimensions where the manifold is undersampled.

Linear ID estimation methods are based on the spectrum of the covariance matrix of the points. Let  $\lambda_k$ ,  $k = 1, \dots, N$  be the eigenvalues of the covariance matrix in decreasing order ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ ). The corresponding eigenvectors are commonly termed *principal components (PCs)* [56]. The ID can be determined by assessing the ‘relevant number’ of PCs. The most common method consists in searching for a linear subspace accounting for a large fraction of the total variance  $\mathcal{V} = \sum_{k=1}^N \lambda_k$ , i.e., identifying the minimum  $K$  such that

$$ID_{PCA,\alpha} = \min K : \frac{\sum_{k=1}^K \lambda_k}{\mathcal{V}} \geq \alpha \quad (1)$$

(typical choices are  $\alpha = 0.8, 0.9, 0.95, 0.99$ ). Another method computes the *participation ratio* [57]

$$ID_{PR} = \frac{1}{\sum_{k=1}^N (\lambda_k/\mathcal{V})^2} \quad (2)$$

$ID_{PR}$  corresponds to an ‘effective number’ of relevant dimensions, as typically  $ID_{PR} \simeq e^{H(\lambda)}$  with  $H(\lambda) = \sum_{k=1}^N (\lambda_k/\mathcal{V}) \log(\lambda_k/\mathcal{V})$  the entropy of the covariance matrix spectrum. Finally, *parallel analysis (PA)* [58] determines the number of significant PCs through a null model. Surrogate data are constructed by resampling the original data by shuffling each data coordinate independently. For each surrogate dataset, the eigenvalues of the covariance matrix are computed. By considering a large number of shuffles, one obtains a null distribution of eigenvalues. Let  $\nu_\alpha$  be the  $(1 - \alpha) \cdot 100$ -th percentile of the null distribution (typically,  $\alpha = 0.05$ ). One finally considers as significant all original eigenvalues exceeding this critical value,

$$ID_{PA} = \max K : \lambda_K \geq \nu_\alpha \quad (3)$$

Linear methods are appropriate only when the manifold is flat, i.e., it lies on a hyperplane. In the presence of curvature, they can largely overestimate the ID.

Among non-linear methods, the most commonly used are *fractal* [59,60] and *nearest-neighbor* methods [45–47]. Fractal methods are based on the fact that the number of points found within a spherical volume of radius  $r$  follows an ID-dependent scaling law. The most common example is the Correlation Dimension (CorrDim) method first proposed in the classic paper by Grassberger and Procaccia [59]. CorrDim focuses on the *correlation integral* (the density of points within a given cutoff distance  $r$ ):

$$\rho_X(r) = \frac{2}{P(P-1)} \sum_{1 \leq i < j \leq P} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (4)$$

If the points are sampled from a distribution on a  $D$ -dimensional manifold, then  $\rho_X(r) \sim r^D$  for  $r \rightarrow 0$ . Thus, one can extract the ID by measuring the slope of the linear fit of  $\rho$  as a function of  $r$  in a log-log scale ( $D = \lim_{r \rightarrow 0} \log \rho_X(r) / \log r$ ). This method can work reasonably well for non-linear manifolds, but is fragile with respect to undersampling, as it requires a number of points scaling exponentially with the ID (“curse of dimensionality”), and thus underestimates high IDs. Nearest-neighbor methods are based on the fact that distances between nearest neighbors obey statistical relations that depend parametrically on the ID. The max-likelihood estimator [45] assumes that the data are distributed on a  $D$ -dimensional manifold and that the density is uniform in a neighborhood containing the first  $K$  neighbors of each point. Under this assumption, consider  $\xi_i = \frac{1}{D} \sum_{k=1}^{K-1} \log(\frac{r_{i,k}}{r_{i,k+1}})$  where  $r_{i,k}$  are the distances of the  $k$ -th neighbors of point  $\mathbf{x}_i$  in the data. One can show that

$$\xi_i \sim \text{Gamma}(k-1, 1) \tag{5}$$

from which one can estimate  $D$  with a maximum likelihood approach on the empirical  $\xi_i$ . The two-nearest-neighbor (Two-NN) estimator [47] restricts the uniform density assumption to the neighborhood containing only the first 2 neighbors of each point. Under this assumption, one can show that

$$\mu_i = \frac{r_{i,2}}{r_{i,1}} \sim \text{Pareto}(D) \tag{6}$$

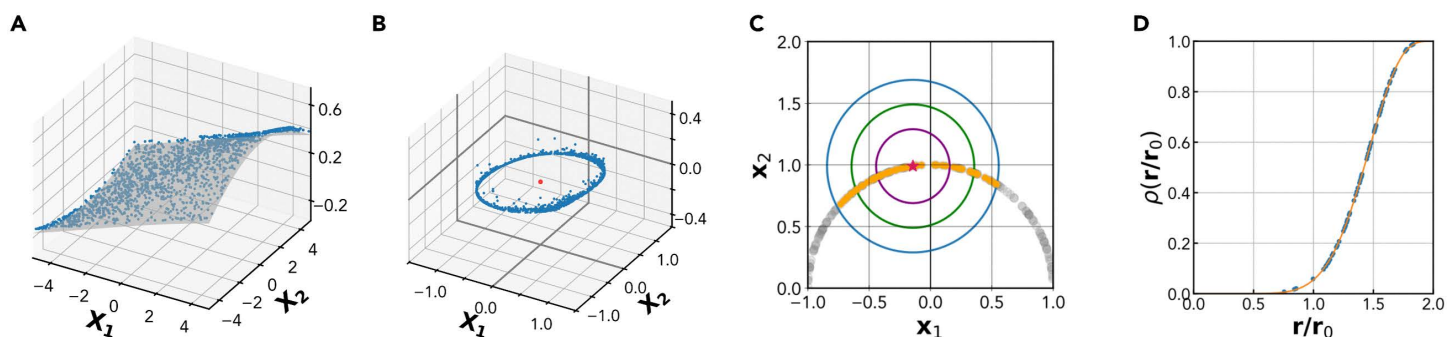
and one can infer  $D$  through a non-linear fit of the empirical distribution of  $\mu_i$ . Also nearest neighbor methods incur the curse of dimensionality.

### The FCI estimator

The full correlation integral (FCI) estimator proposed by Erba et al. [48] was shown to be remarkably robust to under-sampling, being able to correctly estimate large IDs even in the extreme case  $P < D$ . The starting point of the method is provided by CorrDim. To address the curse of dimensionality, Erba et al. made the more restrictive assumption that points are sampled from a rotationally-invariant probability distribution on a  $D$ -dimensional hyperplane. In this case, after normalization of each point to unit norm, the data must lie on a  $D - 1$ -dimensional hypersphere (e.g., a circle for  $D=2$ , a spherical surface for  $D=3$ ), and one can exactly calculate the average correlation integral as a function of  $r$ :

$$\mathbb{E}_x[\rho_x(r)] = \frac{1}{2} + \frac{\Omega_{D-1}}{2\Omega_D}(r^2 - 2) {}_2F_1\left(\frac{1}{2}, \frac{1 - \frac{D}{2}}{\frac{3}{2}} \parallel (r^2 - 2)^2\right) \tag{7}$$

where  ${}_2F_1$  is the (2,1)-hypergeometric function and  $\Omega_D$  is the solid angle in dimension  $D$ . To stress that this analytic formula holds away from the small radius limit of CorrDim, the average correlation integral is named *Full Correlation Integral (FCI)*. With this definition, the FCI, as a function of  $r$ , has a sigmoidal shape with a slope that depends on  $D$ . Performing a non-linear regression on the empirical density of neighbors Eq (4) using the FCI Eq (7) results in an algorithm to determine the ID, shown in Fig 1. In detail, the steps are:



**Fig 1. The Full Correlation Integral (FCI) method.** Panels A-D illustrate the basic steps of the FCI method for ID estimation. **A** The original  $D$ -dimensional dataset **B** Centering and normalization of the dataset: the data points are projected on an unitary sphere in  $D - 1$  dimensions. **C** The empirical correlation integral of the dataset,  $\rho(r/r_0)$ , is measured as a function of the normalized radius  $r/r_0$ . This amounts to computing how many pairs of points are within a distance  $r$ . **D** ID estimation:  $\rho(r)$  is fitted with its theoretical model given by Eq (7), which depends parametrically on the ID. To obtain the true ID, the estimated ID is increased by one, compensating the removal of one degree of freedom during the normalization step.

<https://doi.org/10.1371/journal.pcbi.1014162.g001>

1. *Center and normalize the data* (Fig 1A,1B). Calculate the center of mass  $\mathbf{b}$  of the empirical data as  $\mathbf{b} = 1/P \sum_{i=1}^P \mathbf{x}_i$  and subtract this quantity from each data point, obtaining a data set centered at the origin. Normalize each sample as  $\mathbf{x}_i = \mathbf{x}_i / \|\mathbf{x}_i\|$ .
2. *Fit the FCI*. Measure the empirical correlation integral of the dataset at different radii  $r$  (Fig 1C) and perform a non-linear regression of this empirical neighbors density using the average FCI in Eq (7) as the nonlinear model, and  $D$  as the free parameter (Fig 1D); increase the estimate by one,  $D \rightarrow D + 1$  to reinstate the degree of freedom removed during the normalization step.

Strong deviations from the predicted functional form of the FCI can be detected by the goodness of fit (GoF), measured as the root-mean-squared deviation of the empirical FCI from its best fit using Eq (7) (in S2 Fig, we report examples of fits with different values of GoF).

By construction, the FCI method is exact for linearly embedded Euclidean spaces sampled with a rotationally-invariant probability distribution. Erba et al. showed that the method is remarkably robust with respect to violations of the rotational invariance assumption. In particular, they showed that FCI estimates work very well even in some cases of strongly non-isotropic distributions (e.g., points sampled from corners of  $D$ -dimensional hypercubes). In addition, estimates are very robust with respect to undersampling, as correct estimates of  $D \sim 200$  can be obtained from as little as  $P \sim 20$  points.

### Recurrent neural networks performing cog-tasks

RNNs have become a customary tool of computational neuroscience research [61–64]. We trained RNNs to solve the ‘cog-task’ battery first introduced by Yang et al. [49]. For simplicity, we maintained the same RNN architecture and dynamics as originally proposed.

The network has  $N_{rec} = 256$  recurrent units in the hidden layer coupled with  $N_{in} = 65$  input units and  $N_{out} = 33$  output units. The input layer is composed of one fixation unit and two rings of 32 units representing two stimuli. The output layer has one fixation unit and a ring of 32 units representing motor output.

In the rate model of network activity, each neuron is represented by a continuous variable  $\mathbf{r}(t) \in \mathbb{R}^{N_{rec}}$  representing the neuron’s membrane potential. The dynamics are described by the equation:

$$\tau \frac{d\mathbf{r}}{dt} = -\mathbf{r} + f \left( W^{rec} \mathbf{r} + W^{in} \mathbf{u} + \mathbf{b} + \sqrt{2\tau\sigma_{rec}^2} \boldsymbol{\xi} \right) \quad (8)$$

Here,  $\tau = 100$  ms represents the neuronal time constant, reflecting the slow synaptic dynamics driven by the NMDA receptors. The variable  $\mathbf{u} \in \mathbb{R}^{N_{in}}$  denotes input to the network, while  $\mathbf{b} \in \mathbb{R}^{N_{rec}}$  is the bias or background input. The function  $f(\cdot)$  captures the neuronal nonlinearity, and  $\boldsymbol{\xi}$  represents  $N_{rec}$  independent Gaussian white noise processes with zero mean and unit variance, and  $\sigma_{rec} = 0.05$  quantifies the noise intensity. A typical Softplus function is used,  $f(x) = \log(1 + \exp(x))$ , which after re-parametrization is very similar to the neuronal nonlinearity (frequency-current curve) commonly used in neural circuit models. A set of output units  $\mathbf{z}$  read out nonlinearly from the network,

$$\mathbf{z} = g(W^{out} \mathbf{r}) \quad (9)$$

where  $g(x) = 1/(1 + \exp(-x))$  is the logistic function that limits output activities between 0 and 1.  $W^{in}$ ,  $W^{rec}$ ,  $W^{out}$  are the input, recurrent, and output connection matrices, respectively. After using the first-order Euler approximation with a time-discretization step  $\Delta t$ , we have

$$\mathbf{r}_t = (1 - \alpha) \mathbf{r}_{t-1} + \alpha f \left( W^{rec} \mathbf{r}_{t-1} + W^n \mathbf{u}_t + \mathbf{b} + \sqrt{2\alpha^{-1} \sigma_{rec}^2} \mathbf{N}(0, 1) \right) \quad (10)$$

Here,  $\alpha \equiv \Delta t / \tau$ , and  $N(0, 1)$  represent the standard normal distribution. We use a discretization step  $\Delta t = 20$  ms. We imposed no constraint on the sign or structure of the weight matrices  $W^n$ ,  $W^{rec}$ ,  $W^{out}$ . The network and training were implemented in TensorFlow [65]. The network receives three types of noisy input,

$$\begin{aligned} \mathbf{u} &= (u_{fix}, \mathbf{u}_{mod1}, \mathbf{u}_{mod2}) + \mathbf{u}_{noise} \\ \mathbf{u}_{noise} &= \sqrt{2/\alpha} \sigma_{in} \mathbf{N}(0, 1) \end{aligned} \quad (11)$$

The input noise strength is set to  $\sigma_{in} = 0.01$ . The fixation input  $u_{fix}$  is usually at the high value of 1 when the network is expected to maintain fixation, and changes to 0 when the network is supposed to react. The stimulus inputs  $\mathbf{u}_{mod1}$  and  $\mathbf{u}_{mod2}$  represent one-dimensional circular variables characterized by the angle around a circle. This mirrors what happens in typical experiments with primates, involving *directional stimuli and responses* (for instance, the stimulus is a cloud of dots moving in a given direction, and the animal is asked to move the arm in the same direction). There are 32 units in each of the corresponding stimulus rings, with their preferred directions evenly distributed from 0 to  $2\pi$  ( $\psi_i = 2\pi \cdot i/32 = \pi/16 \cdot i$ ). For unit  $i$ , which has a preferred direction  $\psi_i$ , its activity in response to a stimulus at direction  $\psi$  is given by

$$u_i = \gamma \cdot 0.8 \exp \left[ -\frac{1}{2} \left( \frac{8 |\psi - \psi_i|}{\pi} \right)^2 \right], \quad (12)$$

where  $\gamma$  represents the stimulus intensity. For multiple stimuli, input activities are summed.

The tasks used here are the same as those used by Yang et al. [49]. They selected 20 interrelated tasks, representing common tasks used in neurophysiological studies on nonhuman animals, and useful to understanding basic cognitive processes such as memory-guided response, simple perceptual decision making (DM), context-dependent DM, multisensory integration, parametric working memory, inhibitory control, delayed match-to-sample and delayed match-to-category tasks. The tasks can be divided into three families: the Go, Decision Making (DM), and Matching families.

**Go family.** In this family, a single stimulus is randomly shown either in modality 1 or 2, and the response should be in the same direction of the stimulus. This family includes the forward-Go (Fd-Go), in which the network should respond when the fixation cue goes off; the reaction time Go (RT-Go), in which the network should respond as soon as the stimulus appears; the delay go (Dly-Go), in which the stimulus is turned off before the go and the network has to respond only after the fixation cue goes off. In the ‘anti’ versions of the tasks (Fd-Anti, RT-Anti and Dly-Anti), the network should respond in a direction opposite to that of the stimulus.

**DM family.** In this family, two stimuli are shown simultaneously and are presented till the end of the task. A stimulus is drawn randomly on the whole circle while the other is drawn uniformly between  $90^\circ$  and  $270^\circ$  away from the other. The network has to respond in the strongest of the two directions. The family includes DM1 and DM2 where the two stimuli are presented only in modality 1 (respectively, 2); Ctx DM1 and Ctx DM2 in which the network has to consider only the two stimuli presented in modality 1(2) ignoring those in the other modality; MultiSensory DM in which the network has to respond in the direction of the stimulus that has a stronger combined strength in modalities 1 and 2. In the delayed versions of the tasks (Dly DM 1, Dly DM 2, Ctx Dly DM 1 and Ctx Dly DM 2), the two stimuli are not contiguous in time. They are both shown briefly and are separated by a delay period. Another short delay period follows the offset of the second stimulus before the response.

**Matching family.** In this family, two stimuli are presented consecutively in modality 1 or 2 and separated by a delay period. The response depends on whether or not the two stimuli “match”. This family includes the delayed-match-to-sample

(DMS), delayed-non-match-to-sample DNMS, delayed-match-to-category (DMC) and delayed-non-match-to-category DNMC tasks. In DMS and DNMS, two stimuli match if they point towards the same direction; in DMC and DNMC, if their direction belongs to the same 'category', the first category ranging from 0 to 180°, the other from 180 to 360°. In DMS and DMC the network has to respond towards the second stimulus when there is a match, and fixate otherwise; in DNMC and DNMS, it has to respond towards the second stimulus when the two stimuli are not matched, and fixate otherwise.

### Synthetic neural recordings

Altan et al. [31] created synthetic data to investigate the performance of ID estimators on neural manifolds. In particular, they constructed synthetic data of known dimensionality, mimicking the properties of multi-electrode neural activity recorded in the primary motor cortex of macaques. Data generation starts by extracting  $d \times M$  samples from an empirical distribution of firing rates obtained from multi-electrode array recordings of the macaque primary motor cortex [66] with a binning of 50 ms. Samples are extracted randomly across all recorded neurons and time bins, resulting in temporally uncorrelated samples. The  $d$ -dimensional data are then temporally smoothed (with a Gaussian kernel of s.d. 50 ms) and multiplied by a  $N \times d$  mixing matrix with entries randomly selected from a Gaussian distribution of zero mean and unit variance. This produces a dataset  $X$  composed of  $M$  samples of dimension  $N$ . This embedding step does not change the intrinsic dimension  $d$ . The value of  $d$  ranges from 3 to 40, while  $N=96$  is similar to the number of electrodes in multi-electrode measurements.

In order to produce a dataset with strongly non-linear properties, an exponential nonlinearity is computed on each of the  $N$  variables in  $X$ , yielding a data set

$$Y = f(X) = \frac{e^{\alpha X} - 1}{e^{\alpha} - 1} \quad (13)$$

where  $\alpha$  determines the degree of nonlinearity. Also this non-linear embedding procedure preserves the ID. However, since the transformation is not an isometry (it does not preserve distances), it may significantly distort neighborhood relations (i.e., points that are near before the transformation may be mapped to points that are far apart), and for a finite sampling retrieving the correct ID may be significantly hard.

### Responses to visual stimuli in mouse V1

We reanalyzed the data from [22], which were made available through Figshare [67]. For detailed methods we refer to the original study [22]. Here we provide a brief summary of the salient elements. *Animals*. While the dataset includes data from 8 mice, here we restricted attention to one animal for which responses to several stimulus sets were available. The mouse ('M170714\_MP032') was bred to express tdTomato in inhibitory neurons, and GCaMP6s was expressed virally, identifying excitatory neurons by lack of tdTomato expression. *Imaging* In each of 4 separate recordings, neural activity was recorded through a two-photon microscope using multi-plane acquisition with planes 30 to 35  $\mu m$  apart; 10 or 12 planes were acquired sequentially, scanning the entire stack repeatedly at 3 Hz or 2.5 Hz. The mouse was free to run on an air-floating ball; a field of view was selected such that  $\sim 10,000$  neurons could be observed, with clear calcium transients and a retinotopic location. In total,  $10,075 \pm 1,453$  (mean  $\pm$  s.d.) neurons were obtained in the recordings. *Stimuli*. We analyzed four separate recordings. In two recordings, stimuli consisted of different normalized images from the ImageNet database [68], with 2,800 different images used in total (the researchers manually selected images that had a mix of low and high spatial frequencies and that did not consist of more than 50% uniform background. All images were uniformly contrast-normalized by subtracting the local mean brightness and dividing by the local mean contrast). The 2,800 natural image stimuli were displayed twice in a recording in two blocks of the same randomized order. All stimuli were presented for 0.5 s with a random

interstimulus interval between 0.3 and 1.1 s consisting of a gray-screen. In one recording, a smaller set of 32 images from the same set was presented in a randomized order 114 times, with similar presentation parameters. Finally, in one recording drifting gratings of 32 directions, spaced evenly at  $11.25^\circ$ , were presented 117 times each, lasting 0.5 s each, and with a gray-screen inter-stimulus interval between 0.3 and 1.1 s. The spatial frequency of the stimuli was 0.05 cycles per degree and their temporal frequency was 2 Hz. *Preprocessing.* Calcium movie data were processed using the Suite2p toolbox to estimate spike rates of neurons. Underlying neural activity was estimated using nonnegative spike deconvolution. These deconvolved traces were normalized to the mean and SD of their activity during a 30-min period of gray-screen spontaneous activity. For further detail, please see the original study. Stimulus responses were computed from the first two frames acquired after stimulus onset for each plane. All analyses done in this paper were performed on the preprocessed data available on figshare [67].

### ID and dimensionality reduction

Once an estimate of the ID is achieved, it can be coupled with a dimensionality reduction method to obtain a lower-dimensional representation of the data. For nonlinear manifolds, there are several well-established methods to do so. For instance, Locally Linear Embedding (LLE) [69] assumes the data lie on a smooth manifold of dimension  $D$  and that each point can be approximated by a linear combination of its nearest neighbors. For each data point  $\mathbf{x}_i$ , LLE computes reconstruction weights  $w_{ij}$  by minimizing

$$\min_{w_{ij}} \left\| \mathbf{x}_i - \sum_{j \in \mathcal{N}(i)} w_{ij} \mathbf{x}_j \right\|^2 \quad \text{subject to} \quad \sum_j w_{ij} = 1$$

These weights capture the local geometry of the data. LLE then finds a  $D$ -dimensional embedding  $\mathbf{y}_i$  that preserves these relationships by minimizing

$$\min_{\mathbf{y}_i} \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|^2$$

which leads to an eigenvalue problem. The resulting embedding preserves local neighborhood structure. Isomap [70] is another common nonlinear dimensionality reduction technique, similar to LLE in spirit. Isomap first uses a graph representation of the data to approximate geodesic distances on a manifold. It then applies classical multidimensional scaling on the geodesic distances to project the data in  $D$  dimensions. Recently, an extension of Isomap working for high-dimensional data sets has been proposed [71]. An alternative to classical methods is offered by deep autoencoders [72] - neural networks that learn compact, low-dimensional representations of data by training an encoder-decoder architecture to reconstruct the input. By stacking multiple nonlinear layers, they can capture complex, hierarchical structures that other methods cannot.

## Results

### The local FCI pipeline

We start by presenting our pipeline for intrinsic dimension (ID) estimation. Our method builds on the Full Correlation Integral (FCI) estimator [48] (Methods). FCI assumes an isotropic point distribution on a hyperplane centered on the center-of-mass of the points. While FCI is exceptionally robust in high dimension compared to other methods, it can give suboptimal performance when its assumptions are violated, i.e., in the presence of curvature or violations of the isotropy assumptions. Ref [48] provided preliminary evidence that both limitations could be addressed by performing ID estimates locally on selected neighborhoods on the manifold. Indeed, i) local neighborhoods approximate the tangent plane to the manifold

and are approximately flat and ii) the local distribution of points is typically approximately isotropic, unless a very strong density gradient is present. However, Ref [48] did not turn this into a systematic procedure for local ID estimation. Here, we leverage such insight to propose a systematic and robust procedure for ID estimation.

The main limitation of FCI can be illustrated with a simple example. When points are sampled from a uniform distribution on a two-dimensional plane, FCI yields a very accurate estimate,  $D=1.97$  (Fig 2A). When points are uniformly sampled on a Swiss roll - a curved two-dimensional surface embedded in a three-dimensional space - FCI overestimates the dimension, yielding  $D=2.73$  (Fig 2E). Let us try to address this limitation by exploiting local ID estimates. For any given point  $\mathbf{x}_i$  in the dataset, one can consider its first  $K$  neighbors,

$$\mathcal{B}_K(\mathbf{x}_i) = \{\mathbf{x}_j \text{ such that } \|\mathbf{x}_j - \mathbf{x}_i\| < r_{i,K}\}$$

where  $r_{i,K}$  is the distance of the  $K$ -th neighbor. Local neighborhoods of different size are shown for the plane in Fig 2B. By restricting the ID estimation to points within the neighborhood  $\mathcal{B}_K(\mathbf{x}_i)$ , one obtains a local ID estimate  $D_{i,K}$ . One can randomly subselect  $M \gg 1$  points and consider neighborhoods of increasing size  $K$  centered at each of these points, computing  $D_{i,K}$  for each point and neighborhood size. Estimates  $D_{i,K}$  as a function of the scale  $K$  can be collectively organized in a *multiscale ID plot*. This is shown for the 2-D plane in Fig 2C: estimates at all scales give values in the narrow range [1.7,2.2]. By collecting all local estimates for different  $K$  in a single *local ID histogram*, we obtain a sharp peak around the expected value, 2 (Fig 2D).

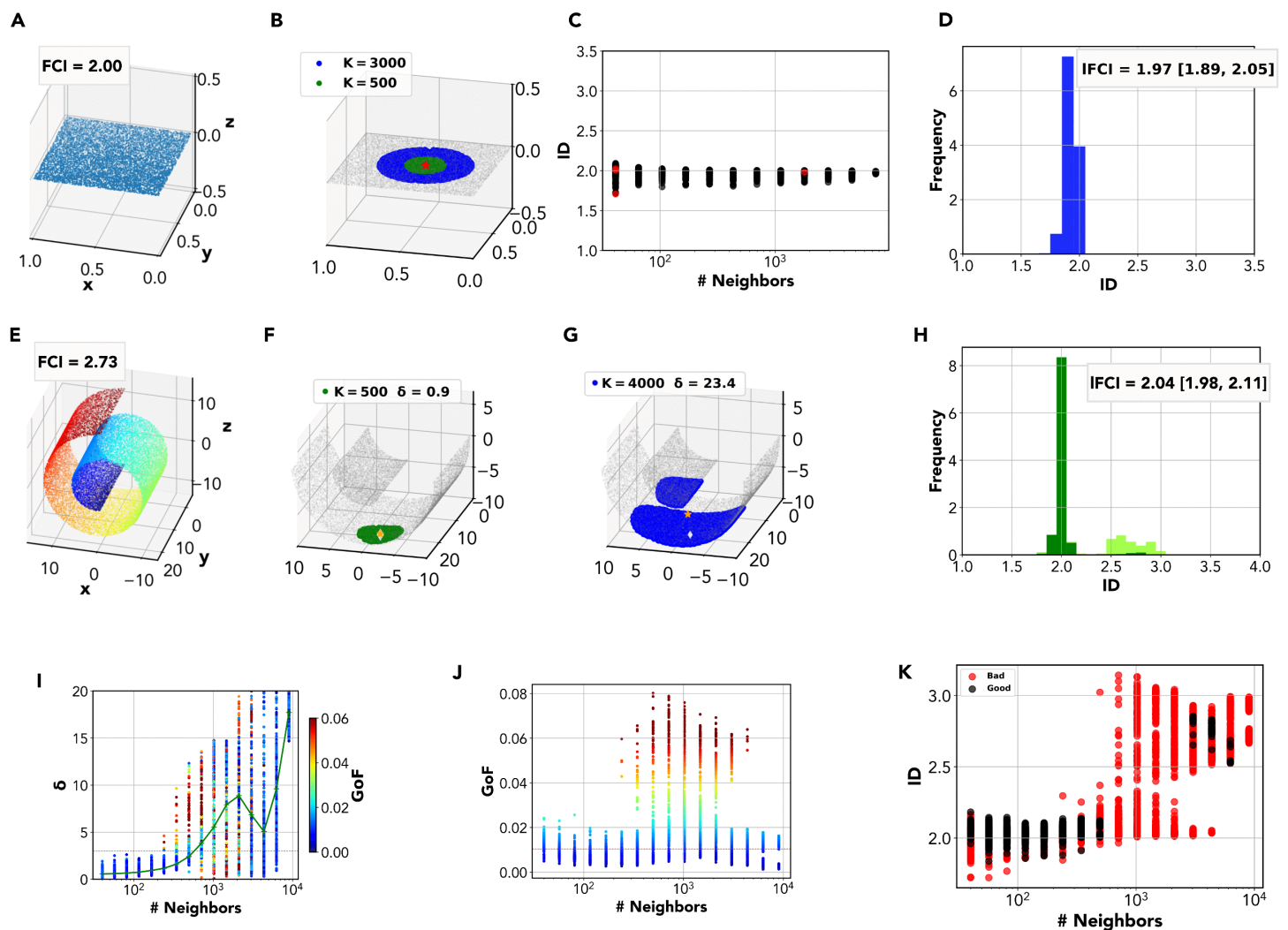
When replicating this procedure on the Swiss roll, we observe that estimates  $D_{i,K}$  remain around 2 for small neighborhood sizes ( $K \lesssim 10^3$ ), while they progressively approach the value 3 for larger neighborhoods (Fig 2H). This is rather unsurprising: for sufficiently small  $K$ , the neighborhoods remain nearly flat, approximating the local tangent plane of the manifold, and local ID estimates correspond to the manifold's proper ID; for larger  $K$ , the neighborhoods become curved and the estimates converge to the standard global FCI estimate. This suggests that *a proper ID estimate could be obtained from the collection of local estimates, removing curved neighborhoods from the analysis*.

Two metrics can track the departure of local neighborhoods from the tangent plane. The first is the position of the neighborhood's *center-of-mass*,  $\mathbf{b}_{i,K} = \frac{1}{K} \sum_j \mathbf{x}_j$ ,  $\mathbf{x}_j \in \mathcal{B}_K(\mathbf{x}_i)$ . For a flat neighborhood, the center-of-mass  $\mathbf{b}_{i,K}$  lies among the points of the neighborhood: as a result, the center-of-mass's distance from the closest point in the neighborhood is of the same order as the typical distance between the nearest neighbors within the neighborhood. For a curved neighborhood, instead,  $\mathbf{b}_{i,K}$  is displaced from the neighborhood (for an extreme case, think of a spherical surface: its center-of-mass is the sphere's center, located far off all surface points). This behavior is effectively captured by the metric

$$\delta_{i,K} = \frac{\min_j \|\mathbf{b}_{i,K} - \mathbf{x}_j\|}{\frac{1}{K} \sum_j \min_k \|\mathbf{x}_j - \mathbf{x}_k\|} \quad \{\mathbf{x}_j, \mathbf{x}_k \in \mathcal{B}_K(\mathbf{x}_i)\} \quad (14)$$

which is simply the distance between the center-of-mass and the closest neighborhood point, normalized by the average nearest-neighbor distance in  $\mathcal{B}_{i,K}$ . For flat neighborhoods, we expect  $\delta \sim 1$ , while curved neighborhoods will give  $\delta > 1$ . The second metric is the FCI Goodness-of-Fit (GoF) (Methods), measured as the average discrepancy between the empirical FCI curve and its fit (i.e., larger values correspond to worse fit quality). Because the center-of-mass is located off the manifold, the points are not distributed isotropically around it, as assumed by the FCI model. Therefore, the fit quality is often poor.

The effectiveness of these metrics can be exemplified by considering two neighborhoods of different sizes on the Swiss Roll. In Fig 2F, we show a neighborhood with  $K=500$ . The neighborhood approximates the tangent plane at  $\mathbf{x}_i$  and it is nearly flat, yielding  $\delta_{i,K} = 0.9$ . Correspondingly, we obtain a good fit (GoF=0.003) and an accurate ID estimate  $D_{i,K}=2.04$ . In Fig 2G, we show a larger neighborhood with  $K=4000$ . The neighborhood strongly deviates from the tangent plane at  $\mathbf{x}_i$ ,



**Fig 2. The Local FCI method.** **A** A dataset with uniform distribution on a plane,  $ID=2$ . FCI correctly identifies the ID of the dataset, yielding  $ID_{FCI}=2.00$ . **B** Local ID estimates can be computed using FCI on neighborhoods of increasing size  $K$ . **C** In the *multiscale ID plot*, local ID estimates are shown as function of the neighborhood size  $K$ . Estimates are colored according to their reliability, as assessed by Goodness-of-Fit (GoF). **D** All reliable local ID estimates for different  $K$  are collected in the *local ID histogram*. The mode of local ID estimates (the peak of the histogram) yields a single overall ID estimate. We call this ID estimation method the *local FCI estimation* (IFCI) method. For the 2-D plane, we obtain  $ID_{IFCI}=1.97$  [1.89,2.05]. **E** A dataset with points uniformly distributed on a Swiss Roll. Even though the ID is 2, FCI gives  $ID_{FCI}=2.73$ . The global FCI overestimates the ID due to the curvature of the manifold. **F** Neighborhoods of small size ( $K \lesssim 10^3$  approximate well the local tangent plane to the manifold. The center-of-mass is located within the neighborhood (low  $\delta$ ). **G** As the neighborhood size increases, neighborhoods depart from the local tangent plane, becoming curved and even possibly disconnected. As a result, the center-of-mass is located outside of the neighborhood (high  $\delta$ ). **H** The local ID histogram for the Swiss Roll shows a prominent peak at 2. We obtain  $ID_{IFCI}=2.04$  [1.98, 2.11]. **I** Distribution of the ratio between distance of the center-of-mass and the average distance between points,  $\delta$ , for the different neighborhoods sizes. The value of  $\delta$  increases at the increasing of the neighborhood size and, around  $K \gtrsim 10^3$ , most of the neighborhoods become curved with  $\delta > 2$  ( $\delta = 2$  is the red line in the plot). In green the median value of  $\delta$ . **J** Distribution of the GoF values for the different neighborhoods sizes. The red line is the threshold chosen as the smallest 99<sup>th</sup> percentile of the GoF distributions for the different neighborhoods sizes.  $GoF_{thr}=0.01$ . The neighborhoods with size  $K \sim 10^3$  are the less reliable. **K** In the *multiscale ID plot* for the Swiss Roll, we observe a sharp transition: for  $K \lesssim 10^3$ , local ID estimates are close to 2, while for larger  $K$  they approach 3. However, we can discriminate bona fide estimates ( $\delta < 2$  and  $GoF < 0.01$ ), shown in black, from unreliable ones ( $\delta > 2$  or  $GoF > 0.01$ ), shown in red. For  $K \gtrsim 10^3$ , most estimates are unreliable.

<https://doi.org/10.1371/journal.pcbi.1014162.g002>

yielding  $\delta_{i,K} \sim 15$ . Correspondingly, we obtain a much worse fit (GoF=0.022) and an excessive ID estimate  $D_{i,K}=2.65$ . The just exemplified behavior is systematic, as shown in Fig 2J. For small neighborhoods ( $K \lesssim 500$ ) we invariably obtain  $\delta \sim 1$ . For larger ones ( $K \gtrsim 500$ ),  $\delta$  progressively grows and the fit quality degrades.

Extensive tests on data sampled on flat manifolds showed that  $\delta$  remains in the range [0,2] with near certainty (S1 Text and S1 Text and S8 Fig). Therefore,  $\delta > 2$  is a robust and conservative criterion for curvature detection, and we can consider local ID estimates as unreliable (due to curvature) if  $\delta > 2$ . The range of GoF values, instead, can widely vary within a dataset, even a flat one. Neighborhoods of specific sizes  $K$  can yield much worse GoF than others not only because they are more curved, as in the case of the Swiss Roll, but also because isotropy violations become more prominent (S1 Text and S9 Fig). When this wide variability occurs, in the sake of accuracy it is reasonable to restrict attention to the best local estimates. To this aim, we may identify a ‘reference’ GoF distribution by considering the size  $K$  at which the best (lowest) GoF is observed, and consider local estimates as unreliable (due to comparatively bad fit quality) if the corresponding GoF lies out of this distribution. In the multiscale ID plot for the Swiss Roll, (Fig 2K) we display reliable estimates in black and unreliable ones in red. Nearly all reliable ID estimates are around 2; conversely, excessive ( $\sim 3$ ) estimates are nearly always spotted as unreliable. Congruently, while the local ID histogram is bimodal, with a peak at 2 (determined by small neighborhoods) and a peak at  $\sim 3$  (determined by large neighborhoods), excluding unreliable estimates suppresses the peak at 3 (Fig 2H).

We thus propose the following *local FCI (IFCI)* pipeline:

1. compute local estimates  $D_{i,K}$  on randomly selected neighborhoods with different values of  $K$ .
2. identify the scale  $\bar{K}$  yielding estimates with the lowest GoF, and compute the 99% percentile  $\gamma$  of the GoF distribution for neighborhoods of size  $K$ .
3. Discard estimates corresponding to sizable curvature ( $\delta_{i,K} > 2$ ) or comparatively poor fit quality (GoF  $> \gamma$ )
4. Collect all local ID estimates in a histogram and take its highest peak  $D^*$  as the best estimate of the overall ID.
5. Specify a plausible range for ID by reporting the 10th and 90th percentile of the local ID estimates.

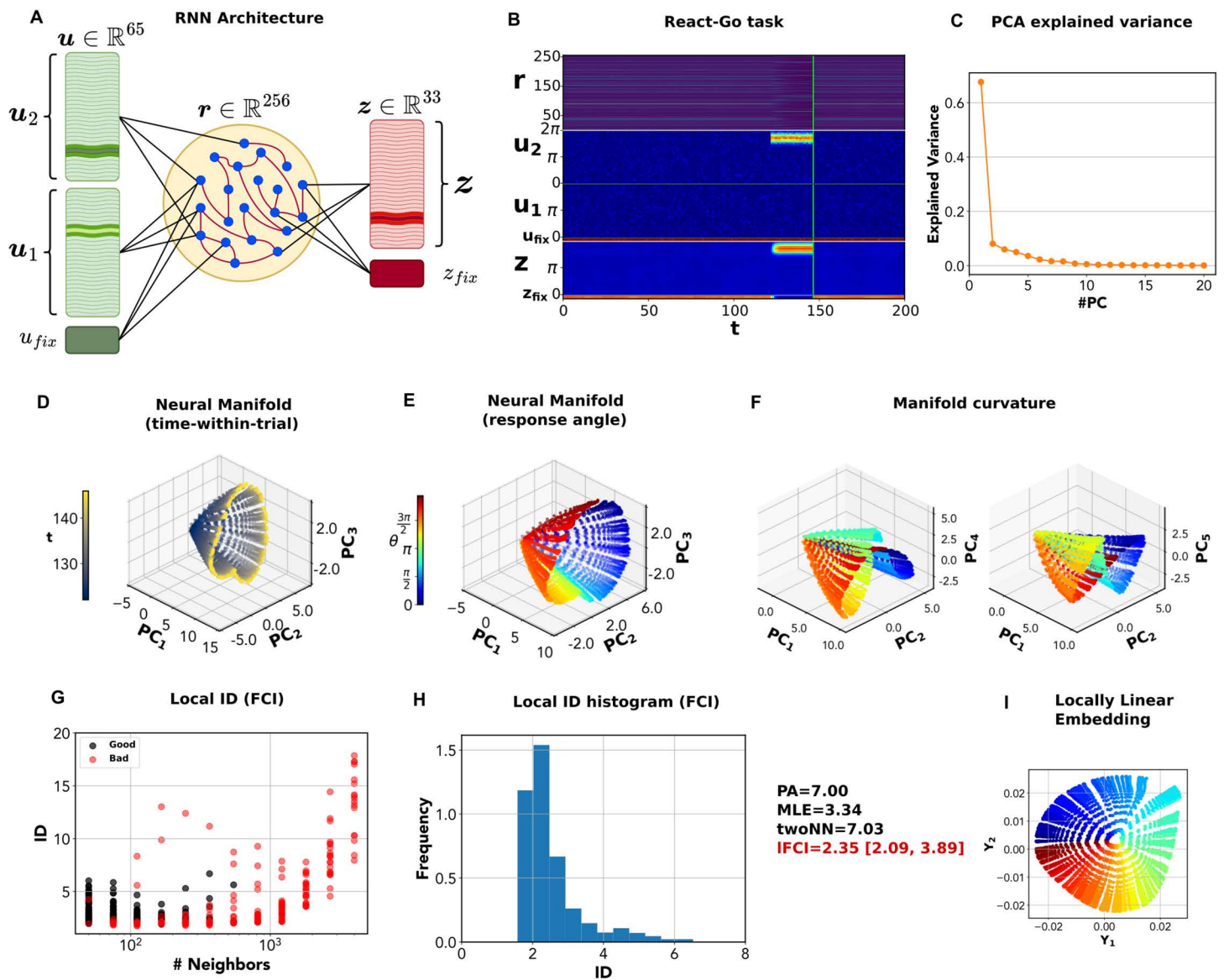
This pipeline is expected to work well for sufficiently well-sampled Riemannian manifolds, such that local neighborhoods will approximate an isotropic sampling on the manifold’s tangent plane.

The computational complexity of IFCI scales as  $\mathcal{O}(N^2DM)$  (S1 Text). As  $N$  and  $D$  are fixed by the data set, the major free parameter entering the computational cost is the number of centers,  $M$ . For our results, we used  $M=100$  centers, which ensures an approximate covering of the whole dataset even when using small neighborhoods ( $K \sim 100$ ). Running the code on a Desktop workstation with 8 Intel i7-8550U cores for a dataset of 10000 points took  $\sim 5$  minutes.

## ID of neural manifolds of recurrent neural networks performing the cog-task battery

We trained recurrent neural networks (RNNs) to perform the ‘cog-tasks’ battery [49] (Fig 3A). For each task, we trained  $M=10$  networks. After training, each RNN performed  $N=200$  trials of the task. For each network, we defined the neural manifold as the set of configurations (in the space of activities of the recurrent layer neurons) visited by the network across the  $N$  trials. We characterized each manifold with common ID estimators.

Before analyzing how the manifold’s ID varies across tasks and trainings, we provide a detailed analysis of an exemplary case: the neural manifold of a RNN performing the React-Go task, where the network must respond with an output in the same direction of the input stimulus (Fig 3B). We first performed principal component analysis. The first principal component explains 64% of the variance, while four subsequent ones explain 5–10% (Fig 3C). As a result, parallel analysis finds  $ID_{PA}=7$ . In the space defined by the first 3 principal components, the manifold appears to be topologically equivalent to a cone (Fig 3D and 3E). The cone axis is roughly aligned to the first principal component, while for fixed values of the



**Fig 3. Intrinsic dimension of the neural manifold in the React-Go task.** **A** Schematic structure of the RNN used for the Cog-Task battery. The network has 65 input units: two sets of 32 channels each,  $u_1$ ,  $u_2$ , respectively encode two angular stimuli that jointly determine the network's response; the fixation input  $u_{fix}$  instructs the network on when to respond. The recurrent layer has 256 units. The output layer has 33 units: 32 units represent an (angular) motor response,  $z$ , while the last unit represents a 'fixation' response  $z_{fix}$ . **B** Input, output and network activity for the RT-Go task. In this task, a single angular stimulus is presented, and the network should immediately give a response in the same direction. **C** Variance explained by successive principal components (PCs) of the network activity. **D** Projection of the network activity on the first three principal components. The color code stands for time-within-trial (time from trial onset).  $PC_1$  is nearly aligned with time-within-trial. Starting from the origin (null activity), the trajectories reach a 'ring' of attractor points (yellow). **E** Projection of the network activity on the first three principal components. The color code stands for response angle. At any time-within-trial, responses lie on a 'ring' encoding the response angle. **F** Projection onto different principal components. The manifold is not entirely contained in the space spanned by the first 3 principal components, but 'stretched' along several orthogonal axes. **G** The multiscale ID plot shows ID estimates as a function of  $K$ , in black (reliable estimates) or red (unreliable ones). At high  $K > 500$ , all estimates become unreliable. **H** The local ID histogram shows a narrow peak around 2. Estimating the mode of the distribution yields  $ID_{FCI} = 2.35 [2.09, 3.89]$ . A linear ID estimation method, parallel analysis (PA), finds  $ID_{PA} = 7$ . The classical geometric maximum-likelihood estimator (MLE) yields  $ID_{MLE} = 3.34$ , while Two-NN yields  $ID_{TwoNN} = 7.03$ , a severe overestimation due to its high sensitivity to local noise. **I** We used locally linear embedding (LLE) to find a global low-dimensional representation of the neural manifold for RT-Go. LLE flattens the manifold onto a circle spanned by a radial coordinate, representing time-within-trial, and an angular coordinate, representing response angle.

<https://doi.org/10.1371/journal.pcbi.1014162.g003>

first principal component, the points describe a (deformed) circle in the space spanned by the second and third principal component. In Fig 3D, points are colored according to time-within-trial ( $t=0$  start of trial;  $t=152$  end of trial); in Fig 3E, according to response angle. These representations allow a ready explanation for the manifold's shape. Before the stimulus presentation, the network is silent ( $\mathbf{r}=0$ ). This configuration corresponds to the vertex of the cone. As the stimulus is presented, the network reacts by reaching a set of attractor states that depend on the response angle and approximately describe a circle. These attractor points are not reached immediately, hence trajectories span a 'conical' surface. Importantly, the surface does not lie exactly in the linear space spanned by the first three principal components. In fact, the cone is 'tilted' in several dimensions. This is shown in Fig 3F, where we replace the third principal component with successive principal components. As a result, linear estimates of the ID end up to be considerably larger than 3.

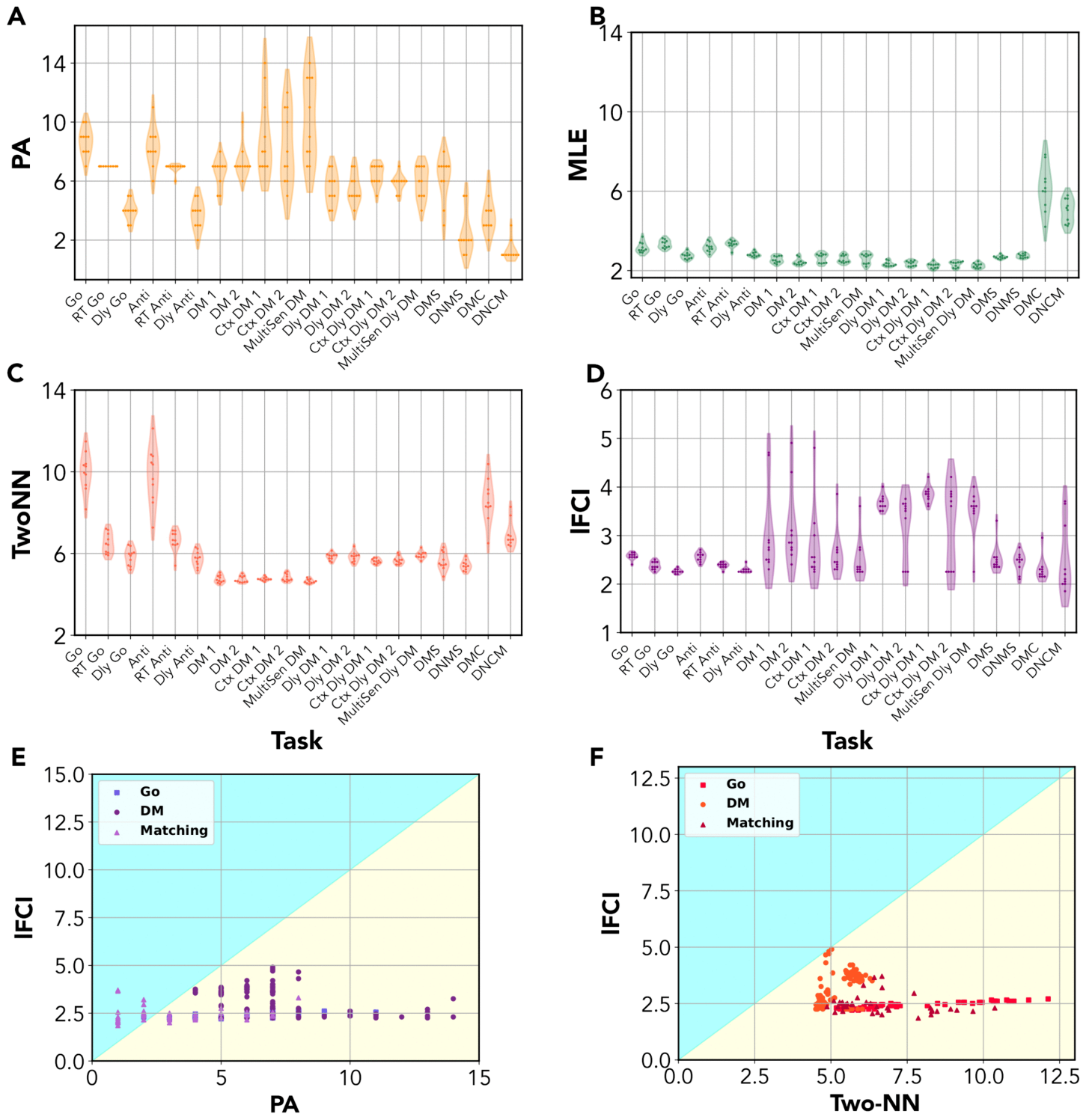
In principle, the 'conical' surface could be described by an axial coordinate (encoding time-within-trial) and an angular one (encoding response angle). Thus, we would expect  $ID=2$ . The best estimate by IFCI was  $ID_{IFCI}=2.35$  (C.I. [2.09, 3.89]). A closer inspection through the multiscale ID plot (in 3G) shows that local ID estimates are low for sufficiently small neighborhoods ( $K \lesssim 1000$ ), while they become much higher for larger neighborhoods, eventually reaching values in the range [10, 20]. Yet, estimates obtained for  $K \gtrsim 500$  are discarded due to curvature and/or bad GoF. As a result, the local ID histogram has a clear peak around 2 (Fig 3H). The intrinsically 2-dimensional geometry of the manifold can be effectively captured using locally linear embedding (Methods), which yields an explicit 2-dimensional parametrization of the manifold. As shown in Fig 3I, the manifold can be 'flattened' on a two dimensional surface parametrized by a radial (time) and an angular (response angle) coordinate. However, the classical maximum likelihood ID estimator yields  $ID_{MLE}=3.34$  - considerably lower than parallel analysis, but definitely larger than 2. The Two-NN estimator gives  $ID_{TwoNN}=7.03$ . As is well known, the maximum-likelihood estimator can be affected by the non-uniformity of the distribution, in this case probably causing an overestimate. The Two-NN estimate, in turn, is strongly biased by small-scale noise [47]. Only by reducing the density of the dataset through decimation (which reduces the effects of small-scale noise increasing the average distance between points), we obtained an estimate  $ID_{TwoNN}=2.05$  much more consistent with the IFCI one (S3 Text and S4 Fig).

In summary, even in this simple task, the neural manifold is highly curved and tilted in several dimensions. As a result, linear ID estimation methods considerably overestimate the dimension. On the contrary, IFCI yields an estimate that is congruent with the task structure and the possibility of parametrizing the manifold with two coordinates.

In Fig 4, we show ID estimates obtained when training 10 RNNs to solve each of the tasks. As for IFCI, we show the best estimate. For all tasks in the Go Family, we consistently obtain  $2.2 \leq ID_{IFCI} \leq 2.6$  (Fig 4D). The classical maximum-likelihood estimator gave ID estimates between 3 and 4 (Fig 4B). Two-NN gave larger and quite variable ID estimates,  $ID_{TwoNN} \in [5, 12]$  (Fig 4C). When applying decimation, which removes small-scale noise, we obtained estimates consistent with those of IFCI, i.e., around 2.5 (S3 Text and S4D and S4F Fig). This result is consistent with the above analysis of the react-Go task, showing that the manifold is essentially spanned by a circular and a temporal variable. Linear ID estimates by parallel analysis are in general much larger (up to  $ID_{PA}=10$ ) and strongly depend on the task and training instance ( $ID_{PA} \in [6, 11]$  for Go and Anti;  $ID_{PA} \in [3, 5]$  for Dly-Go and Dly-Anti;  $ID_{PA}=7$  for RT-Go and RT-Anti). Parallel analysis counts all directions on which the manifold has a non-vanishing projection, which broadly corresponds to the number of principal components explaining 90% of the variance (S4A Fig). Another linear ID estimation method, the participation ratio, effectively counts only the most relevant projections. The participation ratio yields indeed lower estimates than parallel analysis (in the range [1, 5]), but estimates again strongly depend on the task and training instance (S3 Text and S4B Fig).

For the React-Go and React-Anti tasks, the participation ratio can yield estimates close to 1, a direct consequence of the presence of a 'dominant' principal component explaining a large fraction of the total variance (the cone 'axis' representing time-within-trial).

For tasks in the DM family, most trainings gave  $2 \leq ID_{IFCI} \leq 3$  for the non-delayed tasks (with a few outliers in the range [4, 5]), and  $3.5 \leq ID_{IFCI} \leq 4$  for the delayed tasks (with a few outliers around 2). We can trace back this increased dimension to a perturbation of the previously analyzed 'conical' manifold structure. Like in the Go family, neural trajectories



**Fig 4. Intrinsic dimension of neural manifolds for the Cog-Task battery.** For each task in the Cog-Tasks battery, we trained 10 independent recurrent neural networks (RNNs) and computed their ID with different methods. Notably, we compare estimates obtained by the local FCI (IFCI) method with those obtained with a linear ID estimation method, parallel analysis (PA) and two commonly used geometric ID methods, the classical maximum likelihood estimator (MLE) and the two-nearest-neighbors (Two-NN) estimator. **(A)** ID as the number of significant PCs identified with the parallel analysis

(PA). For most tasks, ID estimates are not consistent in different trainings. Most estimates give  $ID_{PA} \geq 5$ . **(B)** ID estimated by the classical maximum likelihood estimator (MLE). MLE gives very consistent estimates, with  $2 \leq ID_{MLE} \leq 3$  for most tasks. The only exception is given by the delayed-match-to-category (DMC) task, which, however, is very difficult to train, with most networks learning the task only approximately. **(C)** For tasks in the 'Go-Anti' family, Two-NN gives quite inconsistent estimates, with  $ID_{TwoNN} \geq 6$ . For tasks in the 'Decision Making (DM)' family, estimates are consistent, with  $ID_{TwoNN} \approx 5$  for the non-delayed and  $ID_{TwoNN} \approx 6$  for the delayed tasks. For tasks in the 'MS' family, we obtain  $ID_{TwoNN} \approx 5$  for D(N)MS, and  $ID_{TwoNN} \geq 6$  for D(N)MC. **(D)** The local FCI (IFCI) method gives consistent estimates around 2.5 for all tasks in the 'Go' family, and estimates between 3 and 4 for tasks in the DM and MS family. **(E)** The ID identified by IFCI is always considerably lower than the ID estimate obtained by PA, except for the MS family. This is a reflection of the fact that neural manifolds for the Cog-Tasks battery are generally curved and extended across many linear dimensions, despite having an intrinsic 2- or 3-dimensional geometry. **(F)** The ID identified by FCI is always lower than the estimate obtained by Two-NN, which is severely affected by small-scale noise.

<https://doi.org/10.1371/journal.pcbi.1014162.g004>

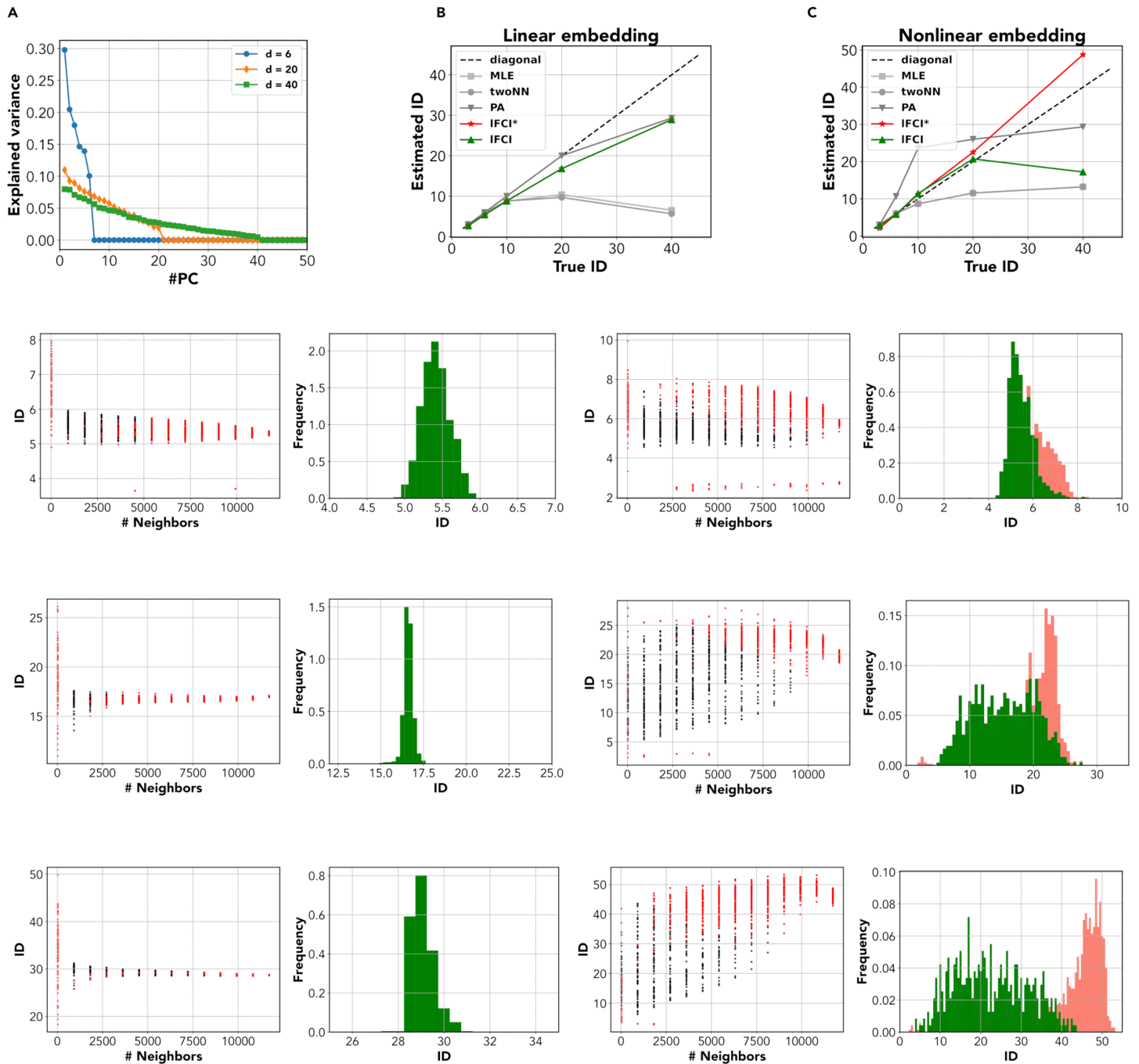
must encode the response angle in the 'Go' phase to produce a correct response. However, in the delayed tasks (where the two stimuli are given in succession) the network must also transiently encode the first stimulus angle, which must be kept in memory until the second stimulus is presented and the correct response can be computed. Trajectories converging to the same attractor points thus form a smeared bundle. This bundle also includes 'backward' trajectories: as the stimuli are switched off in the 'Go' phase, trajectories do not remain in the vicinity of attractor points but slowly converge back to the origin before the end of the trial (for a detailed example and discussion, see [S2 Text](#) and [S3 Fig](#)). This 'smearing' of trajectories associated to the same attractor determines an additional 'transversal' degree of freedom, causing a rise in dimension. This phenomenon also occurs in the non-delayed tasks, where 'smearing' is only due to backwards trajectories not coinciding with the forward ones, causing only a minor dimension increase with respect to the Go tasks ( $\sim +0.5$ ). ID Estimates obtained with classical MLE are consistent and slightly above 2 for all DM tasks ([Fig 4B](#)), suggesting that MLE identifies only the two local dimensions with the strongest variability. ID Estimates obtained with Two-NN are consistently larger than those of IFCI ( $ID_{Two-NN} \simeq 5$  for the non-delayed DM tasks,  $ID_{Two-NN} \simeq 6$  for the delayed ones; [Fig 4C](#)). Upon decimation, estimates are in good agreement with those of IFCI ([S3 Text](#) and [S4D Fig](#)). On the other hand, linear ID estimates are again quite inconsistent over tasks and trainings: parallel analysis generally yields much larger estimates ( $ID_{PA} \in [4, 15]$ ; [S4A Fig](#)), while the participation ratio yields estimates in a similar range ( $2 \leq ID_{PR} \leq 5$ ). The participation ratio estimates are not correlated with those of IFCI ([S3 Text](#) and [S4E Fig](#)), suggesting that the two methods capture different geometrical properties:  $ID_{PR}$  captures global variation along different orthogonal axes, and  $ID_{IFCI}$  captures local variation.

Finally, for tasks of the MS family, we consistently obtain  $ID_{IFCI} \sim 2.5$  for DMS and DNMS. This result is fully consistent with estimates given by maximum-likelihood estimator and Two-NN (after decimation). Parallel analysis gives inconsistent and high estimates, while the participation ratio gives estimates close to 1, due to a dominant PC. For the DMC and DNCM tasks, most trainings give  $2 \leq ID_{IFCI} \sim 2.5$  (except one outlier around 3 for DMC, one outlier slightly below 2 and two outliers between 3 and 4 for DNCM). ID Estimates for DMC are thus more consistent than those of other estimators, while those for DNCM show some variability, which also emerges in maximum-likelihood and Two-NN estimates - perhaps reflecting a difficulty of properly training this task (the accuracy was  $0.90 \pm 0.05$  compared to  $0.97 \pm 0.01$ ).

## ID of synthetic neural recordings

Altan et al. [\[31\]](#) introduced a benchmark data set of synthetic neural recordings. As detailed in Methods, the data were generated by considering firing rate data from  $d$  independent recording channels, with varying  $d \in [3, 6, 10, 20, 40]$ , embedding them in  $N=96$  dimensions with a linear transformation (*linearly embedded data*), and finally applying a non-linear transformation (*nonlinearly embedded data*).

The linearly embedded data present a challenge for classical geometric ID estimators, but not linear ones. While parallel analysis always finds  $ID \simeq d$ , geometric estimators find  $ID \sim 10$  for all  $d \geq 10$ , a reflection of their (known) difficulty in estimating high dimensionalities. Here, IFCI outperforms geometric methods, giving estimates broadly in line with those of parallel analysis:  $ID_{IFCI} = [3.00, 5.63, 8.87, 16.74, 29.41]$  for  $d = [3, 6, 10, 20, 40]$  ([Fig 5B](#)). In fact, this data does not pose



**Fig 5. ID estimation for the synthetic neural data with increasing dimension.** In Ref [31], neural manifolds were constructed by randomly sampling firing rates from  $d$  independent channels in a multielectrode array. Data were then linearly embedded in 96 dimensions (*linearly embedded data*), and further subjected to a non-linear transformation (*nonlinearly embedded data*). **A** Variance explained by the principal components (PCs) of the linearly embedded data. **B** The ID was estimated by different methods, including parallel analysis (PA), Maximum Likelihood Estimator (MLE), Two-Nearest Neighbors (Two-NN), and the local FCI method; in red the IFCI estimates without our procedure to filter out the unreliable estimates, in green the IFCI results after the refinement. Dashed line indicates the identity. The linearly embedded data are stretched along  $d$  orthogonal axes, but the variance is very small ( $\lesssim 10^{-2}$ ) for several components. All ID estimation methods yield  $ID \sim d$  for  $d \leq 10$ . For  $d \geq 10$ , MLE and Two-NN show a non-monotonic behavior, with  $ID \leq 10$ . PA and IFCI give monotonically increasing estimates for increasing  $d$ , with  $ID_{IFCI} \sim ID_{TwoNN} \sim 30$ ; the presence of the unreliable results does not affect the final IFCI estimate. **C** The nonlinearly embedded data give rise to strongly non-smooth manifolds. All ID estimation methods

yield  $ID \sim d$  for  $d \leq 10$ . For  $d \geq 10$ , MLE and Two-NN show a non-monotonic behavior, with  $ID \leq 10$ . IFCI shows a monotonically increasing trend, but the ID remains much lower than  $d$  ( $ID_{IFCI} \leq 13$ ). Keeping the unreliable estimates IFCI gives  $ID_{IFCI} \sim d$  for all  $d$ . **D** For the linearly embedded data at  $d=6$ , local ID estimates are consistent at all  $K$ , with a slightly decreasing trend for increasing  $K$ , and we obtain  $IFCI=5.51$  [5.23, 5.71]. **E** For the nonlinearly embedded data at  $d=6$ , local ID estimates are consistent at all  $K$ , and we obtain  $IFCI=5.51$  [4.98, 6.25]. **F** For the linearly embedded data at  $d=20$ , most local ID estimates are consistently in the range [15, 20] at all  $K$ , with a decreasing trend for increasing  $K$ . We obtain  $IFCI=16.72$  [16.45, 17.06], consistent with the number of PC components with variance  $>10^{-2}$ . **G** For the nonlinearly embedded data at  $d=20$ , local ID estimates are very variable (in the range [5, 20]) for all  $K < 10^4$ . Only for  $K \approx 10^4$  do estimates become more consistent with values around 20. **H** For the linearly embedded data at  $d=40$ , most local ID estimates are consistently in the range [23, 34] at all  $K$ , with a decreasing trend for increasing  $K$ . We obtain  $IFCI=29.33$  [28.77, 29.97], consistent with the number of PC components with variance  $>10^{-2}$ . **I** For the nonlinearly embedded data at  $d=40$ , local ID estimates are very variable (in the range [5, 40]) for all  $K < 10^4$ . Only for  $K \approx 10^4$  do estimates become more consistent with values around 40.

<https://doi.org/10.1371/journal.pcbi.1014162.g005>

significant challenges to IFCI, as reflected in the very well-behaved appearance of the multiscale ID plots and the local ID histograms for  $d=[6, 20, 40]$  (Fig 5D, 5F and 5H). Estimates are reliable (a natural consequence of the non-curved nature of the data) and consistent for all scales  $K$ , so that the local ID histogram is sharply peaked. The ID underestimation (which becomes evident for  $d=20$  and even more so for  $d=40$ ) is due to the presence of many dimensions with extremely small variance,  $<10^{-2}$  (Fig 5A). In this condition, it is essentially impossible to reliably discriminate these dimensions from small-scale noise. In fact, also parallel analysis tends to identify a number of dimensions lower than  $d$ , and  $ID_{IFCI}$  and  $ID_{PA}$  converge for  $d=40$ . To validate this hypothesis, we variance-normalized the data before the embedding. For these equal-variance data, we obtained  $ID_{IFCI} \simeq d$  for all  $d$  (S6 Fig).

The nonlinearly embedded data have some very peculiar characteristics, which we summarize here (only some were discussed in [31]): i) the data distribution along each dimension is nearly log-normal. ii) The upper tails of the distributions give rise to isolated low-density regions along orthogonal axes (S5A Fig). In these regions, the ‘tangent plane’ has not a dimension equal to the ID, and perhaps is not even well-defined. iii) the density varies dramatically across the dataset (to measure this quantitatively, we used the inverse of the distance of the 10-th neighbor as a proxy of the density). As a result, the manifold contains a highly dense ‘core’ and a large, rarefied ‘periphery’ (this is visualized in S5B and S5D Fig, where we projected the manifolds for  $d=20, d=40$  in two dimensions using t-stochastic neighbor embedding [73], showing the local point density).

Due to these joint characteristics, ID estimation becomes a difficult endeavor. Due to i), the covariance structure of the data is strongly influenced by outliers: this makes linear ID estimation unreliable. In fact, parallel analysis first overestimates (for  $d < 20$ ), then underestimates (for  $d=40$ ) the ID. Geometric estimators are bound to underestimate the ID (for  $d > 10$ ) due to their intrinsic limitations but also the manifold geometry, as nearest-neighbor statistics in the peripheral regions strongly deviates from the predicted distribution. Indeed, the maximum-likelihood estimator and Two-NN give  $ID_{MLE}, ID_{Two-NN} < d$  for  $d \geq 10$ . In this situation, IFCI can give accurate estimates up to  $d=20$ , while it fails for  $d=40$ , producing a sharp underestimation:  $ID_{IFCI}=[2.8, 5.8, 11.3, 20.7, 17.2]$  for  $d=[3, 6, 10, 20, 40]$ . The problem for  $d=40$  can be mitigated by relaxing a step in the pipeline, the removal of estimates corresponding to bad GoF, although this comes at the price of some overestimation:  $ID_{IFCI}=[2.5, 5.8, 11.2, 22.5, 48.7]$  for  $d=[3, 6, 10, 20, 40]$ . To better analyze the reasons for this (partial) success, in Fig 5E, 5G and 5I we show the multiscale ID plots and the local ID histograms for  $d=6, 20, 40$ . For small  $d$  ( $d=6$ ), the behavior is similar to the one observed in the linear embedding case. Although local ID estimates are slightly more variable (in the range [5, 7]), the local ID histogram is still sharply peaked. For  $d=20, 40$  local ID estimates appear to be highly variable for most neighborhood sizes  $K$ , except  $K \geq 10000$ . This behavior depends on the dataset geometry: local ID estimates strongly correlate with the density (S5C and S5E Fig). Neighborhoods in the highly dense ‘core’ yield local ID estimates  $\sim d$ , whereas neighborhoods in the rarefied periphery cannot detect the full dimensionality, systematically yielding lower ID estimates. For large  $K$  ( $\gtrsim 5000$ ), all neighborhoods encompass a large fraction of the whole manifold. ID estimates become accordingly more consistent. However, estimates also become unreliable. This is not due to curvature (we always find  $\delta \sim 1$  in our data, reflecting the absence of curvature), but rather to a bad GoF. In

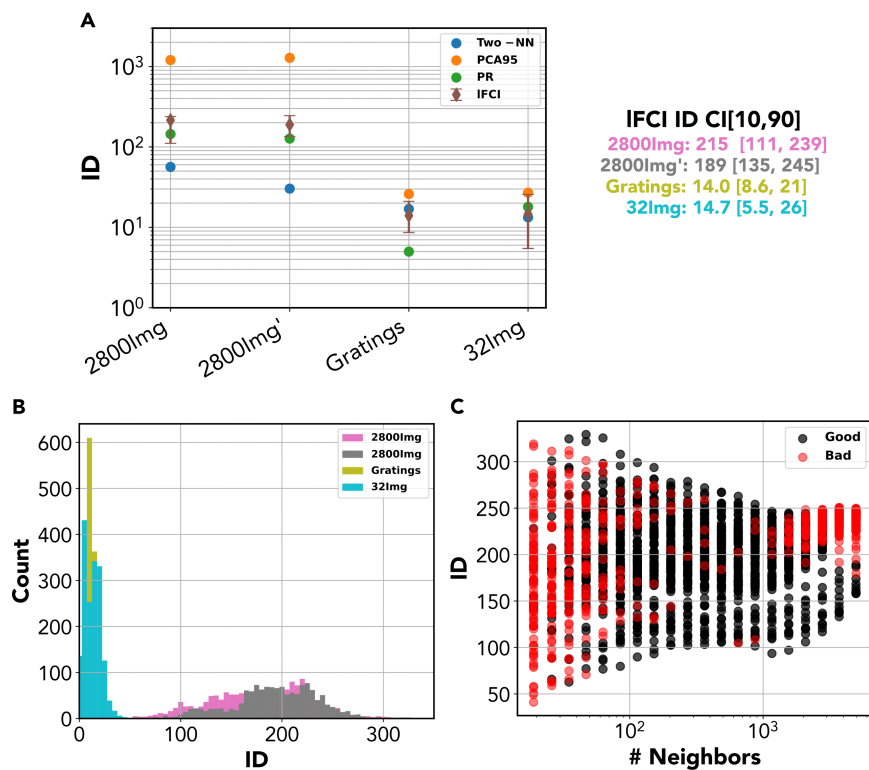
fact, large neighborhood must include regions with variable density, leading to a violation of the isotropy assumption and hence a deviation of the FCI curve from its theoretical shape. The FCI fitting procedure, however, spots a reasonable if slightly overestimated dimension. For  $d=20$ , the local ID histogram has a peak around 20 even if unreliable estimates are discarded, even though the peak is shifted to  $\sim 22$  and becomes more pronounced if they are included (5G). For  $d=40$ , the local ID histogram has a peak around 20 unless unreliable estimates are included (5G), which produces a peak around 45. In synthesis: when considering small neighborhoods, the presence of rarefied regions tends to produce ID underestimation; when using large neighborhoods, this problem is solved but density variations imply a bad GoF and some overestimation. Nevertheless, even for this very challenging manifold, IFCI can yield rich insight into the manifold's structure and ID.

### ID of high-dimensional visual responses

Having extensively benchmarked our method on artificial data in preceding sections, we now analyze real data from 2-photon imaging data of mice exposed to visual stimuli. Data were first discussed in the classical work by Stringer et al. [22,74] and made publicly available [67]. This data set includes responses of  $\sim 10,000$  primary visual cortex neurons to several stimuli, including: 2,800 natural images, 32 natural images, 32 simple grating stimuli, and degraded natural images. Importantly, each visual stimulus was presented at least twice, allowing for a better discrimination of stimulus-related and stimulus-unrelated variability in neural recordings. Ref. [22] characterized the linear ID of neural responses. The ID found was always in line with the size of the stimulus set: for the small stimulus sets (32 images or gratings) it remained of the order of  $\approx 30$ , while for the very high-dimensional stimulus (2800 images), it reached values of the order of  $\sim 10^3$ . Importantly, Ref [22] provided solid evidence that this large linear dimension truly represents stimulus-related variability. Leveraging the availability of (at least) two repetitions of the same stimulus, they computed the cross covariance of responses across repetitions, which is provably a good estimate of stimulus-related variance. They then characterized relevant (linear) dimensions as axes with high cross-covariance (a method called cross-validated PCA or cvPCA, S4 Text).

While Stringer et al.'s findings suggest that visual responses are *not* organized in a low-dimensional hidden manifold, this conclusion cannot be drawn without an analysis of the ID. In fact, a large linear dimension may emerge from a high-ID geometry, but also from a low-ID, highly nonlinear geometry. Note that the latter case is not merely a mathematical curiosity, as some common neural codes were shown to produce exactly this type of highly non-linear manifolds [42], and a recent study showed that a high linear dimension can emerge in recurrent neural networks with a low-dimensional latent dynamics [44]. To adjudicate between these two scenarios, we re-analyzed Stringer et al.'s data using IFCI alongside other ID estimators. We focused our analysis on data from a mouse for whom responses to multiple stimulus sets were recorded. To remove stimulus-unrelated variability, we adopted the same logic as Ref [22], and we first projected the data onto a (linear) subspace containing 99% of stimulus-related variance, as estimated by cvPCA. While this step effectively removes stimulus-unrelated dimensions, it maintains the non-linear geometry of the stimulus-related responses - if any (projecting a nonlinear manifold onto a hyperplane yields a linear manifold only if all the nonlinearity is contained in the dimensions orthogonal to the hyperplane).

In the two experiments with 2,800 images, the linear dimension (as estimated via PCA with  $\alpha = 0.95$ ) is very large,  $\sim 1,200$  (Fig 6A). This is due to many of the linear dimensions having a very low variance: the effective linear dimension (as estimated by the participation ratio) is nearly one order of magnitude lower, around  $\sim 140$ . We estimated the ID using Two-NN and IFCI. Two-NN yields inconsistent estimates in the two replications - respectively,  $\sim 55$  and  $\sim 30$ . While IFCI estimates display a relatively large interval for both repetitions ( $ID_{IFCI} \in [115, 245]$ ,  $ID_{IFCI} \in [111, 239]$ ), the local ID histogram is peaked (Fig 6C) and the best estimates broadly in line with those of the participation ratio,  $ID_{IFCI} = 215$ ,  $ID_{IFCI} = 189$ . To properly interpret this discrepancy, one should take into account that Two-NN tends to systematically underestimate the ID when above  $\sim 20$ , and radically so for  $ID \gtrsim 100$  (S1 Fig) Although it would be tempting to explain the difference between Two-NN estimates and linear dimension estimates in terms of curvature, this explanation



**Fig 6. ID of High-dimensional visual responses.** Results of the ID analysis on the mouse visual responses. The data have been processed with cvPCA before the dimensionality analysis. **A** ID results for the different experiments; **B** Distribution of the IFCI ID values for the different experiments; **C** ID values for the different neighborhood sizes for the experiment with 2800 images. The analysis for the curvature shows that the dataset is flat with  $\delta < 1$  for all the experiments.

<https://doi.org/10.1371/journal.pcbi.1014162.g006>

is not supported by the data. In fact, IFCI failed to detect significant curvature in the data: the distribution of  $\delta$  was entirely contained in the range  $[0, 1]$  for all neighborhood sizes (the 95% percentile of the  $\delta$  distribution was 0.65 in the first replication and 0.65 in the second). The comparatively low estimates given by Two-NN are then probably attributable to a method's bias.

In the experiment with a reduced stimulus set of 32 natural images, both the linear dimension and the ID dropped dramatically with respect to what seen for the full stimulus set. PCA yielded a value  $\sim 28$ , roughly equivalent to the maximum dimension (equal to the size of the stimulus set). Estimates by the other methods were broadly consistent: the participation ratio yields  $\sim 171$ , Two-NN gives  $ID_{Two-NN} \sim 13$ , IFCI  $ID_{IFCI} \sim 20$ . Despite the decrease in dimension, responses do not seem to be arranged in a truly low-dimensional manifold. In the experiment with gratings, results were qualitatively similar, except that we observed a further decrease of the effective linear dimension estimated by the participation ratio, which is  $\sim 5$ . The linear dimension estimated with PCA is  $\sim 27$ , close to the maximum possible value. Two-NN gave  $ID_{Two-NN} \sim 17$ ,  $ID_{IFCI} \sim 12$ .

In summary, our analysis suggests that visual responses to a high-dimensional set of natural images are organized within a high-dimensional, linear manifold with dimension  $\sim 150$ . Even when considering a much smaller stimulus set, or radically simpler stimuli such as grating, there is no evidence of a truly low-dimensional organization. We remark that these conclusions remain valid if one does not preliminarily remove stimulus-unrelated variance (S4 Text and S7 Fig): the main differences are that (i) for the stimulus sets with 2,800 images, the picture is similar but stimulus-unrelated noise

leads to even larger dimension estimates (ii) stimulus-unrelated variability outbalances stimulus-related one in the case of the stimulus sets with 32 stimuli, leading to much larger dimension estimates.

## Discussion

The dimensionality of neural population activity is at the root of key open issues in computational neuroscience, such as structural constraints on the geometry of neural population activity [18,20,75], the relation between spontaneous and task-evoked activity [22,74,76], the development of abstract codes [24,26,77], and the number of recorded neurons needed to exhaustively characterize neural population responses. More generally, characterizing the dimensionality of neural manifolds is critical to understand how neuronal populations collectively encode relevant variables and perform computation (see classic [7] and recent [10] reviews on neural manifolds). The ongoing debate could be affected by the limitation of common methodologies to measure the ID of neural manifolds [17]. Linear methods (PCA and its evolutions) are inappropriate in the case of strongly nonlinear manifolds, while geometric ID estimators (which have been often employed to study representations of deep artificial neural networks [78–81]) can dramatically underestimate the ID in the case of high dimensionality.

## Methodology

Here, we introduced a novel ID estimation pipeline based on a local version of the full-correlation integral (Fig 1). Our pipeline, called local FCI or IFCI, computes the ID locally on neighborhoods of variable size, obtaining a robust ID estimate from the histogram of local estimates (Fig 2). Importantly, we prune the histogram from unreliable local estimates obtained on neighborhoods with significant curvature or density variations. As shown by the summary of results in Table 1, our pipeline is versatile enough to reliably cover a wide range of dimensionalities and non-linearities. In all cases analyzed, IFCI performed comparably or better than other methods. Importantly, no other method (among those included in our analysis) was able to produce consistently reliable results in a wide range of dimensionalities. Our pipeline encountered tangible difficulties only in the extreme case of a high-dimensional, highly skewed data distribution; even in this case, relaxing the pipeline yielded correct insight on the ID (Fig 5). Table 2 shows the IFCI results with the confidence interval. It is worth noticing that the True ID is included in the confidence interval. In the linear embedding case, the ID is underestimated due to the presence of some dimensions that have a very small variance.

**Table 1. Summary of the results.**

Task name	True ID	PA	MLE	twoNN	IFCI
cog-task (GO)	~ 2	6.43	3.09	7.42	<b>2.41</b>
cog-task (DM)	~ 3	7.08	2.23	5.26	<b>3.16</b>
cog-task (Match)	~ 3	3.40	4.14	6.63	<b>2.45</b>
neural traj. (lin)	3	3.00	3.04	2.99	<b>2.70</b>
	6	6.00	5.94	5.88	<b>5.51</b>
	10	10.00	8.82	8.77	<b>8.81</b>
	20	20.00	9.70	10.40	<b>16.72</b>
	40	29.33	5.66	6.54	<b>29.33</b>
neural traj. (exp)	3	3.00	2.98	3.02	<b>2.70 (2.13*)</b>
	6	10.67	6.10	6.04	<b>5.51 (5.77*)</b>
	10	23.67	8.65	8.69	<b>9.31 (11.20*)</b>
	20	26.00	11.58	11.52	<b>16.22 (22.51*)</b>
	40	29.33	13.20	13.24	<b>17.42 (48.72*)</b>

<https://doi.org/10.1371/journal.pcbi.1014162.t001>

**Table 2. Summary of the results for IFCI.**

Task name	True ID	IFCI C.I. [10,90]
neural traj. (lin)	3	2.70 [2.62, 2.96]
	6	5.51 [5.23, 5.71]
	10	8.81 [8.55, 9.08]
	20	16.72 [16.45, 17.06]
	40	29.33 [28.77, 29.97]
neural traj. (exp)	3	2.70 [2.05, 3.22]
	6	5.51 [4.98, 6.25]
	10	9.31 [6.94, 11.39]
	20	16.22 [8.92, 21.87]
	40	17.42 [11.85, 39.73]

<https://doi.org/10.1371/journal.pcbi.1014162.t002>

In future developments, we might integrate this pipeline with suitable denoising procedures to further enhance its robustness. In Ref. [31], Altan et al. proposed the following denoising/estimation scheme: 1) use parallel analysis (PA) to obtain an upper bound on the ID,  $D_U$ ; 2) denoise the data by first projecting (linearly, via PCA, or non-linearly, via an autoencoder) onto a  $D_U$ -dimensional space, and then projecting back into the original space; 3) if nonlinear denoising is more effective than linear denoising (indicating a nonlinear manifold) use a geometric ID estimator; otherwise, use a linear one. Our results show that FCI can yield a more robust upper bound than PA, also valid in a highly nonlinear case. We thus suggest the following modification of Altan et al.'s scheme: 1) use FCI to obtain an upper bound on the ID,  $D_U$ ; 2) denoise the data 3) compute a handful of local ID estimates for small neighborhoods (e.g.,  $M=10$  centers and  $K \in [50, 100, 200]$ ) and average the resulting estimates 4) if there is agreement with the global estimates (small relative error), stop; otherwise, perform to the full IFCI procedure to characterize the ID (note that step 3, which can be performed rapidly at a low computational cost, can spare the computational burden of the complete IFCI procedure when probably not needed).

**'Embedding' vs. 'intrinsic' dimension of neural manifolds.** In Ref [41], Jazayeri et al. explicitly distinguished between the ID and the minimum dimension of a linear subspace containing the manifold, which they termed 'embedding dimension' (ED). Following their terminology, linear methods (such as parallel analysis, participation ratio, and PCA with a fixed variance threshold) essentially measure the ED. Due to their (undeniable) advantages, such as simplicity and robustness, many research works employ linear methods (see, e.g., [14,15,20,26,27,32–37,39]), and many recently proposed developments for dimensionality reduction are also based on linear projections [23,33,38,40]. However, since neural manifolds can be highly non-linear and curved [42], paying attention to the ID/ED distinction may be relevant in many cases. A paradigmatic example is provided by our analysis of neural manifolds of RNNs trained to perform the Cog-task battery (Figs 3 and 4) Albeit intrinsically low-dimensional, these manifolds had non-vanishing projections along several orthogonal axes, and the ED was considerably higher than the ID. Importantly, the ID was i) more clearly related to fundamental task variables such as response angle and integration time ii) much more stable than ED across trainings, suggesting that the ID could be among the 'universal' properties [82] that do not depend on network architecture and training, but only on the computational scaffold underlying the task [82]. In a classic paper, Gao et al. proposed an upper bound to the neural manifold dimension in terms of the 'neuronal task complexity' [36], an index depending on the autocorrelation of activity for similar values of task-relevant variables (such as response angle and integration time). In fact, the neuronal task complexity bound applies to a measure of ED, the participation ratio, and this bound is generally quite loose and much larger than the ID (this is shown in S4 Fig for the 'Cog-Task' battery).

**Low-dimensional neural manifolds for cog-tasks.** Recent literature has abundantly mentioned 'low-dimensional neural manifolds'. In experimental studies, the low-dimensionality was generally inferred from the fact that few

principal components could explain a large fraction of variance (see, e.g., the seminal article by Churchland et al. demonstrating that a reaching task induces a low-dimensional neural activity [15]). However, only rarely was a rigorous ID characterization performed (an exception is [17]). In turn, theoretical studies have often investigated cases where neural activity was compelled to be low-dimensional by structural constraints, such as a low-rank connectivity [18,19,61]. To what extent networks *trained* to perform simple tasks would exhibit a low dimensional activity was therefore an open question. In this work, we have analyzed the activity of RNNs performing the common ‘Cog-Task’ battery, showing that their activity is indeed organized in low-dimensional (approximately 2- or 3-dimensional) manifolds (Fig 4). The low dimensionality seems to be directly related to the simplicity of the tasks at hand. The only key variable the networks must encode is the response angle, which can be encoded by a ‘ring’ of attractors (Fig 3). Additional dimensions come from the need to switch these representations on and off, departing or returning to a state of zero activity. The minimal increase in dimension is + 1, but it can be larger depending on the divergence of forward and backward trajectories. Future studies could investigate more complex tasks such as the ‘Mod-Cog’ battery proposed by [51].

**High dimensional visual responses.** The article by Stringer et al. [22] contained a landmark finding, as it provided robust evidence that neural activity can be very high-dimensional. We reanalyzed the same data to test whether using a non-linear ID estimator would alter, or circumscribe, the main conclusions reached in that work. Our analysis based on IFCI offered substantial backing to the picture drawn by Stringer et al. by showing that visual responses to a large set of visual stimuli are organized in a linear, high-dimensional manifold - and not in a non-linear, low-dimensional one (Fig 6). While the ID is lower than the embedding dimension estimated by PCA, its value ( $\sim 150$ ) aligned with the effective linear dimension corresponding to the most relevant principal components, as estimated by the participation ratio. When analyzing responses to a small random subset of 32 stimuli (from the same set of 2,800 stimuli), the ID was considerably reduced ( $\sim 15$ ), but we found no evidence of truly low-dimensional activity. This is not entirely surprising, because the resulting manifold should emerge from neural configurations that are randomly sampled from the global high-dimensional manifold.

Overall, this analysis, coupled with the previous one on ‘Cog-Tasks’, support the hypothesis that a primary determinant of the neural manifolds’ ID is the complexity of the stimulus-space underlying the observed neural responses.

**Limitations.** The main limitation of the IFCI method is that it assumes a smooth manifold structure. In the presence of non-smooth, sparsely or radically non-uniformly sampled manifolds, IFCI’s assumptions can be violated. Some of the dimensions can be poorly sampled locally, yielding local ID estimates that are significantly lower than the global ID of the manifold. This was partly evidenced in our analysis of nonlinear data from [31], Figs 5 and S5), where low-density, peripheral regions yielded a lower local ID than the high-density bulk of the manifold. As long as real recordings do not present similar strongly skewed, fat-tailed firing rate distributions, IFCI should perform reliably. Related to the first, a second limitation is IFCI’s dependence on the isotropy assumptions. While this assumption is often well-satisfied locally, it can be violated in the presence of strong density gradients, leading to a poor GoF and, possibly, biased estimates. While we decided to stick to a conservative criterion, whereby we discard local estimates with comparatively bad GoF within the same dataset, ideally one could try to gauge the GoF filter, and possibly correct biases, depending on the measured density gradient in the data. Future work may address this point through a wide array of numeric experiments.

Finally, in this work we did not discuss how to obtain low (or lower-) dimensional representations once a solid ID estimate is obtained (we only made a cursory example in the discussion of the neural manifold for the React-Go task). In principle, many approaches for nonlinear dimensionality reduction are available for this aim: from classical methods such as locally linear embedding to more modern ones such as deep autoencoders (see Methods for a brief review). Nearly all of these methods include the ID as a free parameter, so ID estimation would serve as a crucial preliminary step. Future work might address this issue more in depth.

**Conclusion.** In conclusion, we proposed a robust pipeline to estimate the intrinsic dimension of neural manifolds, offering a new tool to tackle an open debate. The inherent properties of IFCI make it robust to undersampling and

nonlinearity, allowing us to probe a large range of IDs. Our pipeline allowed us to rigorously demonstrate that simple tasks induce a low-dimensional structure in neural activity and highlight the difference between the ID and the ‘embedding dimension’ given by linear estimators.

## Supporting information

### S1 Text. Additional methodological remarks on IFCI.

(PDF)

### S2 Text. Intrinsic dimension of neural manifolds for the Cog-Task battery - Neural manifold for Ctx-DM1.

(PDF)

### S3 Text. Intrinsic dimension of neural manifolds for the Cog-Task battery - Additional Results.

(PDF)

### S4 Text. Intrinsic dimension of high-dimensional visual responses - additional notes.

(PDF)

**S1 Fig. Scaling of FCI and Two-NN estimates with increasing D.** We show results of FCI on  $M$  uniformly sampled data ( $M = 10^2, 10^3, 10^4$ ) on hypercubes of known dimension varying from  $D = 10$  to  $D = 1000$ . Results are compared with those of Two-NN, a common ID estimator.

(TIF)

### S2 Fig. Examples of Goodness-of-Fit (GoF) for FCI.

(TIF)

**S3 Fig. Dimensionality of the neural activity manifold in the Ctx-DM1 task.** **A** Input, output and network activity for the Ctx-DM1 task; **B** Input, output and network activity for the Ctx-DM1 task; **C** Explained variance of the different PCs of the network activity. Looking at the variance, PCA predicts a dimensionality of 1 if we consider the highest drop in variance or a dimension of 7 if we consider the second elbow in variance. **D** Projection of the network activity on the first three PCs. The color code stands for time-within-trial (time from trial onset).  $PC_1$  is nearly aligned with time-within-trial. **E** Projection of the network activity on the first three PCs. The color code stands for response angle. **F** Multiscale ID plot for the RT-Go task. At low  $K$ , estimates are close to 3. **G** Local ID histogram, showing a peak at  $D = 3.26$ . **H** Locally linear embedding can flatten the manifold for the Ctx-DM1 task onto a 2-D surface but only with some approximation.

(TIF)

**S4 Fig. Intrinsic dimension of neural manifolds for the Cog-Task battery.** For each task in the Cog-Tasks battery, we trained 10 independent RNNs to solve the task and computed the ID of the neural activity manifolds for each case. **(A)** ID as the number of PCs identified with the principal component analysis (PCA) keeping the first principal components explaining 95% of the variance **(B)** ID estimated by the participation ratio (PR) **(C)** upper bound on the ID given by neuronal task complexity (NTC) **(D)** ID computed through the multiscale (with decimation) Two-NN method. **(E)** NTC can under- and overestimate the ID **(F)** The ID identified by the multiscale Two-NN method and the local FCI are in good agreement.

(TIF)

**S5 Fig. Local ID estimates and density for the ‘multielectrode array’ data.** **(A)** TSNE embedding of the data with  $d = 20$ . Low local IDs are found in the ‘periphery’, while high local IDs are found in the ‘core’ **(B)** local IDs estimates strongly correlate with the local density of points for  $d = 20$ . **(C)** same as (A) for  $d = 40$ . **(D)** local IDs estimates strongly correlate with the local density of points for  $d = 20$ .

(TIF)

**S6 Fig. ID estimation for the ‘multielectrode array’ data with equal variance.**

(TIFF)

**S7 Fig. Analysis of the High-Dimensional Visual Responses.** Results of the raw high dimensional visual responses recorded by Stringer [67]. **A** ID and ED values for the different experiments; **B** IFCI ID distribution for the different experiment. **C** ID as a function of the number of neighbors for the experiment with 2,800 images.

(TIF)

**S8 Fig. The distribution of curvature ( $\delta$ ) for flat manifolds.** Distribution of the values of for three different dataset with increasing anisotropy and fixed  $ID=3$ . **(A)**,  $\alpha = 0$  corresponds to an isotropic distribution. **(B, C)**,  $\alpha = 1$  and  $\alpha = 2$  correspond to non isotropic distributions as points tend to accumulate around the origin and the axes. The data for  $\alpha = 2$  are already strongly non-isotropic as they display a strong density gradient. For each data set, we considered local neighborhoods of varying size  $K$  and computed  $\delta$  and GoF. In panels **D-F**, we show the resulting distributions of  $\delta$  for  $ID=3$ . The distributions are only mildly affected by  $\alpha$  and the values remain strictly bounded. In particular, the 99-th percentile of the  $\delta$  distribution is always lower than  $\delta = 2$  (red dashed line in the plots).

(TIFF)

**S9 Fig. The distribution of GoF for flat manifolds.** **A-C** show the distributions of GoF, grouped according to neighborhood size  $K$ . In this case, the distributions are strongly affected by  $\alpha$ ; in red the distribution with the lowest median. For  $\alpha = 0$  the largest neighborhoods yield the most consistent GoF distribution. For  $\alpha = 1$ , the distribution becomes clearly bimodal for all  $K \geq 1000$ ; For  $\alpha = 2$ , the effects of anisotropy are so strong that most large neighborhoods ( $K \geq 1000$ ) yield  $GoF > 0.01$ . In this case, the neighborhoods yielding the most consistent GoF distribution correspond to  $K \approx 150$ . **(D-F)** ID distributions for the different, in yellow the distribution for all the estimates, in green the distribution of ID after the filtering process. To ensure that analogous results could be obtained in a wide range of dimensions we repeated the same analysis for  $D=3$ ,  $D=10$ ,  $D=20$ ,  $D=40$ . **(G)** the 99-th percentile of the distribution, that always remains below  $\delta = 2$ . **(H)** the 99-th percentile of the GoF distribution corresponding to the  $K$  yielding the lowest median GoF. We see that the GoF threshold is highly variable from dataset to dataset. In all cases, removing ‘bad’ estimates improved the ID estimate **(I)**.

(TIF)

**S10 Fig. ID distribution as a function of the number of centers.** For the Swiss Roll, we computed the local ID histogram varying the number of centers  $M=10,30,100,300,1000$ . The number of centers affects only marginally the distribution of the ID estimates. The size of the neighborhoods is the parameter that most affects the estimation.

(TIF)

## Acknowledgments

We thank Alessandro Laio for precious suggestions and Luca Mazzucato, Samir Suweis, and Bill Bialek for their feedback on the paper.

## Author contributions

**Conceptualization:** Jacopo Fadanni, Pietro Rotondo, Michele Allegra.

**Formal analysis:** Rosalba Pacelli, Pietro Rotondo, Michele Allegra.

**Funding acquisition:** Pietro Rotondo, Michele Allegra.

**Investigation:** Jacopo Fadanni, Rosalba Pacelli.

**Methodology:** Jacopo Fadanni, Rosalba Pacelli, Pietro Rotondo, Michele Allegra.

**Project administration:** Alberto Zucchetta, Michele Allegra.

**Software:** Jacopo Fadanni.

**Supervision:** Alberto Zucchetta, Pietro Rotondo, Michele Allegra.

**Visualization:** Jacopo Fadanni, Rosalba Pacelli.

**Writing – original draft:** Jacopo Fadanni, Michele Allegra.

**Writing – review & editing:** Jacopo Fadanni, Rosalba Pacelli, Pietro Rotondo, Michele Allegra.

## References

1. Urai AE, Doiron B, Leifer AM, Churchland AK. Large-scale neural recordings call for new insights to link brain and behavior. *Nat Neurosci*. 2022;25(1):11–9. <https://doi.org/10.1038/s41593-021-00980-9> PMID: [34980926](https://pubmed.ncbi.nlm.nih.gov/34980926/)
2. Barack DL, Krakauer JW. Two views on the cognitive brain. *Nat Rev Neurosci*. 2021;22(6):359–71. <https://doi.org/10.1038/s41583-021-00448-6> PMID: [33859408](https://pubmed.ncbi.nlm.nih.gov/33859408/)
3. Langdon C, Genkin M, Engel TA. A unifying perspective on neural manifolds and circuits for cognition. *Nat Rev Neurosci*. 2023;24(6):363–77. <https://doi.org/10.1038/s41583-023-00693-x> PMID: [37055616](https://pubmed.ncbi.nlm.nih.gov/37055616/)
4. Chung S, Abbott LF. Neural population geometry: an approach for understanding biological and artificial neural networks. *Curr Opin Neurobiol*. 2021;70:137–44. <https://doi.org/10.1016/j.conb.2021.10.010> PMID: [34801787](https://pubmed.ncbi.nlm.nih.gov/34801787/)
5. Mitchell-Heggs R, Prado S, Gava GP, Go MA, Schultz SR. Neural manifold analysis of brain circuit dynamics in health and disease. *J Comput Neurosci*. 2023;51(1):1–21. <https://doi.org/10.1007/s10827-022-00839-3> PMID: [36522604](https://pubmed.ncbi.nlm.nih.gov/36522604/)
6. Sussillo D. Neural circuits as computational dynamical systems. *Curr Opin Neurobiol*. 2014;25:156–63. <https://doi.org/10.1016/j.conb.2014.01.008> PMID: [24509098](https://pubmed.ncbi.nlm.nih.gov/24509098/)
7. Vyas S, Golub MD, Sussillo D, Shenoy KV. Computation through neural population dynamics. *Annu Rev Neurosci*. 2020;43:249–75. <https://doi.org/10.1146/annurev-neuro-092619-094115> PMID: [32640928](https://pubmed.ncbi.nlm.nih.gov/32640928/)
8. Duncker L, Sahani M. Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings. *Curr Opin Neurobiol*. 2021;70:163–70. <https://doi.org/10.1016/j.conb.2021.10.014> PMID: [34837752](https://pubmed.ncbi.nlm.nih.gov/34837752/)
9. Gort J. Emergence of universal computations through neural manifold dynamics. *Neural Computation*. 2024;36(2):227–70.
10. Perich MG, Narain D, Gallego JA. A neural manifold view of the brain. *Nat Neurosci*. 2025;28(8):1582–97. <https://doi.org/10.1038/s41593-025-02031-z> PMID: [40721675](https://pubmed.ncbi.nlm.nih.gov/40721675/)
11. Anderson PW. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science*. 1972;177(4047):393–6.
12. Mazor O, Laurent G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*. 2005;48(4):661–73. <https://doi.org/10.1016/j.neuron.2005.09.032> PMID: [16301181](https://pubmed.ncbi.nlm.nih.gov/16301181/)
13. Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat Neurosci*. 2009;12(7):919–26. <https://doi.org/10.1038/nn.2337> PMID: [19483687](https://pubmed.ncbi.nlm.nih.gov/19483687/)
14. Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 2013;503(7474):78–84. <https://doi.org/10.1038/nature12742> PMID: [24201281](https://pubmed.ncbi.nlm.nih.gov/24201281/)
15. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, et al. Neural population dynamics during reaching. *Nature*. 2012;487(7405):51–6. <https://doi.org/10.1038/nature11129> PMID: [22722855](https://pubmed.ncbi.nlm.nih.gov/22722855/)
16. Gallego JA, Perich MG, Miller LE, Solla SA. Neural manifolds for the control of movement. *Neuron*. 2017;94(5):978–84. <https://doi.org/10.1016/j.neuron.2017.05.025> PMID: [28595054](https://pubmed.ncbi.nlm.nih.gov/28595054/)
17. Altan E, Ma X, Miller LE, Perreault EJ, Solla SA. Low-dimensional neural manifolds for the control of constrained and unconstrained movements. *bioRxiv*. 2023;2023:2023–05.
18. Mastrogiuseppe F, Ostojic S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*. 2018;99(3):609–23.e29.
19. Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S. The role of population structure in computations through neural dynamics. *Nat Neurosci*. 2022;25(6):783–94. <https://doi.org/10.1038/s41593-022-01088-4> PMID: [35668174](https://pubmed.ncbi.nlm.nih.gov/35668174/)
20. Cimeša L, Ciric L, Ostojic S. Geometry of population activity in spiking networks with low-rank structure. *PLoS Comput Biol*. 2023;19(8):e1011315. <https://doi.org/10.1371/journal.pcbi.1011315> PMID: [37549194](https://pubmed.ncbi.nlm.nih.gov/37549194/)
21. Pezon L, Schmutz V, Gerstner W. Linking neural manifolds to circuit structure in recurrent networks. *bioRxiv*. 2024. 2024–02.
22. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. *Nature*. 2019;571(7765):361–5. <https://doi.org/10.1038/s41586-019-1346-5> PMID: [31243367](https://pubmed.ncbi.nlm.nih.gov/31243367/)
23. Manley J, Lu S, Barber K, Demas J, Kim H, Meyer D, et al. Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron*. 2024;112(10):1694–1709.e5. <https://doi.org/10.1016/j.neuron.2024.02.011> PMID: [38452763](https://pubmed.ncbi.nlm.nih.gov/38452763/)
24. Bernardi S, Benna MK, Rigotti M, Munuera J, Fusi S, Salzman CD. The Geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*. 2020;183(4):954–967.e21. <https://doi.org/10.1016/j.cell.2020.09.031> PMID: [33058757](https://pubmed.ncbi.nlm.nih.gov/33058757/)

25. Ganguli S, Sompolinsky H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu Rev Neurosci.* 2012;35:485–508. <https://doi.org/10.1146/annurev-neuro-062111-150410> PMID: [22483042](https://pubmed.ncbi.nlm.nih.gov/22483042/)
26. Fusi S, Miller EK, Rigotti M. Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol.* 2016;37:66–74. <https://doi.org/10.1016/j.conb.2016.01.010> PMID: [26851755](https://pubmed.ncbi.nlm.nih.gov/26851755/)
27. Elmoznino E, Bonner MF. High-performing neural network models of visual cortex benefit from high latent dimensionality. *PLoS Comput Biol.* 2024;20(1):e1011792. <https://doi.org/10.1371/journal.pcbi.1011792> PMID: [38198504](https://pubmed.ncbi.nlm.nih.gov/38198504/)
28. Srinath R, Ni AM, Marucci C, Cohen MR, Brainard DH. Orthogonal neural representations support perceptual judgements of natural stimuli. *bioRxiv.* 2024.
29. Li Q, Sorscher B, Sompolinsky H. Representations and generalization in artificial and brain neural networks. *Proc Natl Acad Sci U S A.* 2024;121(27):e2311805121. <https://doi.org/10.1073/pnas.2311805121> PMID: [38913896](https://pubmed.ncbi.nlm.nih.gov/38913896/)
30. Kuoch M, Chou CN, Parthasarathy N, Dapello J, DiCarlo JJ, Sompolinsky H. Probing biological and artificial neural networks with task-dependent neural manifolds. In: *Conference on Parsimony and Learning.* 2024. p. 395–418.
31. Altan E, Solla SA, Miller LE, Perreault EJ. Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *PLoS Comput Biol.* 2021;17(11):e1008591. <https://doi.org/10.1371/journal.pcbi.1008591> PMID: [34843461](https://pubmed.ncbi.nlm.nih.gov/34843461/)
32. Clark DG, Abbott LF, Litwin-Kumar A. Dimension of activity in random neural networks. *Phys Rev Lett.* 2023;131(11):118401. <https://doi.org/10.1103/PhysRevLett.131.118401> PMID: [37774280](https://pubmed.ncbi.nlm.nih.gov/37774280/)
33. Dahmen D, Recanatani S, Ocker GK, Jia X, Helias M, Shea-Brown E. Strong coupling and local control of dimensionality across brain areas. *Biorxiv.* 2020;2020:2020–11.
34. Fascianelli V, Battista A, Stefanini F, Tsujimoto S, Genovesio A, Fusi S. Neural representational geometries reflect behavioral differences in monkeys and recurrent neural networks. *Nat Commun.* 2024;15(1):6479. <https://doi.org/10.1038/s41467-024-50503-w> PMID: [39090091](https://pubmed.ncbi.nlm.nih.gov/39090091/)
35. Gao P, Ganguli S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr Opin Neurobiol.* 2015;32:148–55. <https://doi.org/10.1016/j.conb.2015.04.003> PMID: [25932978](https://pubmed.ncbi.nlm.nih.gov/25932978/)
36. Gao P, Trautmann E, Yu B, Santhanam G, Ryu S, Shenoy K. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv.* 2017. <https://doi.org/10.1101/214262>
37. Parthasarathy A, Herikstad R, Bong JH, Medina FS, Libedinsky C, Yen S-C. Mixed selectivity morphs population codes in prefrontal cortex. *Nat Neurosci.* 2017;20(12):1770–9. <https://doi.org/10.1038/s41593-017-0003-2> PMID: [29184197](https://pubmed.ncbi.nlm.nih.gov/29184197/)
38. Pellegrino A, Stein H, Cayco-Gajic NA. Dimensionality reduction beyond neural subspaces with slice tensor component analysis. *Nat Neurosci.* 2024;27(6):1199–210. <https://doi.org/10.1038/s41593-024-01626-2> PMID: [38710876](https://pubmed.ncbi.nlm.nih.gov/38710876/)
39. Perkins SM, Cunningham JP, Wang Q, Churchland MM. Simple decoding of behavior from a complicated neural manifold. *BioRxiv.* 2023;2023:2023-04.
40. Zimnik AJ, Ames KC, An X, Driscoll L, Lara AH, Russo AA. Identifying interpretable latent factors with sparse component analysis. *bioRxiv.* 2024.
41. Jazayeri M, Ostojic S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr Opin Neurobiol.* 2021;70:113–20. <https://doi.org/10.1016/j.conb.2021.08.002> PMID: [34537579](https://pubmed.ncbi.nlm.nih.gov/34537579/)
42. De A, Chaudhuri R. Common population codes produce extremely nonlinear neural manifolds. *Proc Natl Acad Sci U S A.* 2023;120(39):e2305853120. <https://doi.org/10.1073/pnas.2305853120> PMID: [37733742](https://pubmed.ncbi.nlm.nih.gov/37733742/)
43. Fortunato C, Bannasar-Vázquez J, Park J, Chang JC, Miller LE, Dudman JT. Nonlinear manifolds underlie neural population activity during behaviour. *bioRxiv.* 2023.
44. Schmutz V, Haydaroglu A, Wang S, Feng Y, Carandini M, Harris KD. High-dimensional neuronal activity from low-dimensional latent dynamics: a solvable model. *bioRxiv.* 2025.
45. Levina E, Bickel P. Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information Processing Systems.* 2004;17.
46. Campadelli P, Casiraghi E, Ceruti C, Rozza A. Intrinsic dimension estimation: relevant techniques and a benchmark framework. *Mathematical Problems in Engineering.* 2015;2015(1):759567.
47. Facco E, d'Errico M, Rodriguez A, Laio A. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci Rep.* 2017;7(1):12140. <https://doi.org/10.1038/s41598-017-11873-y> PMID: [28939866](https://pubmed.ncbi.nlm.nih.gov/28939866/)
48. Erba V, Gherardi M, Rotondo P. Intrinsic dimension estimation for locally undersampled data. *Sci Rep.* 2019;9(1):17133. <https://doi.org/10.1038/s41598-019-53549-9> PMID: [31748557](https://pubmed.ncbi.nlm.nih.gov/31748557/)
49. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci.* 2019;22(2):297–306. <https://doi.org/10.1038/s41593-018-0310-2> PMID: [30643294](https://pubmed.ncbi.nlm.nih.gov/30643294/)
50. Driscoll LN, Shenoy K, Sussillo D. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat Neurosci.* 2024;27(7):1349–63. <https://doi.org/10.1038/s41593-024-01668-6> PMID: [38982201](https://pubmed.ncbi.nlm.nih.gov/38982201/)
51. Khona M, Chandra S, Ma JJ, Fiete IR. Winning the lottery with neural connectivity constraints: faster learning across cognitive tasks with spatially constrained sparse RNNs. *Neural Comput.* 2023;35(11):1850–69. [https://doi.org/10.1162/neco\\_a\\_01613](https://doi.org/10.1162/neco_a_01613) PMID: [37725708](https://pubmed.ncbi.nlm.nih.gov/37725708/)
52. Flesch T, Nagy DG, Saxe A, Summerfield C. Modelling continual learning in humans with Hebbian context gating and exponentially decaying task signals. *PLoS Comput Biol.* 2023;19(1):e1010808. <https://doi.org/10.1371/journal.pcbi.1010808> PMID: [36656823](https://pubmed.ncbi.nlm.nih.gov/36656823/)

53. Riveland R, Pouget A. A neural model of task compositionality with natural language instructions. *bioRxiv*. 2022.
54. Masse NY, Rosen MC, Tsao DY, Freedman DJ. Flexible cognition in context-modulated reservoir networks. *bioRxiv*. 2022;2022:2022–05.
55. Camastra F. Data dimensionality estimation methods: a survey. *Pattern Recognition*. 2003;36(12):2945–54. [https://doi.org/10.1016/s0031-3203\(03\)00176-6](https://doi.org/10.1016/s0031-3203(03)00176-6)
56. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202. <https://doi.org/10.1098/rsta.2015.0202> PMID: [26953178](https://pubmed.ncbi.nlm.nih.gov/26953178/)
57. Kramer B, MacKinnon A. Localization: theory and experiment. *Rep Prog Phys*. 1993;56(12):1469–564. <https://doi.org/10.1088/0034-4885/56/12/001>
58. Buja A, Eyuboglu N. Remarks on parallel analysis. *Multivariate Behav Res*. 1992;27(4):509–40. [https://doi.org/10.1207/s15327906mbr2704\\_2](https://doi.org/10.1207/s15327906mbr2704_2) PMID: [26811132](https://pubmed.ncbi.nlm.nih.gov/26811132/)
59. Grassberger P, Procaccia I. Characterization of strange attractors. *Phys Rev Lett*. 1983;50(5):346–9. <https://doi.org/10.1103/physrevlett.50.346>
60. Kégl B. Intrinsic dimension estimation using packing numbers. *Advances in Neural Information Processing Systems*. 2002;15.
61. Sussillo D, Barak O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput*. 2013;25(3):626–49. [https://doi.org/10.1162/NECO\\_a\\_00409](https://doi.org/10.1162/NECO_a_00409) PMID: [23272922](https://pubmed.ncbi.nlm.nih.gov/23272922/)
62. Barak O. Recurrent neural networks as versatile tools of neuroscience research. *Curr Opin Neurobiol*. 2017;46:1–6. <https://doi.org/10.1016/j.conb.2017.06.003> PMID: [28668365](https://pubmed.ncbi.nlm.nih.gov/28668365/)
63. Barrett DG, Morcos AS, Macke JH. Analyzing biological and artificial neural networks: challenges with opportunities for synergy?. *Current Opinion in Neurobiology*. 2019;55:55–64.
64. Yang GR, Molano-Mazón M. Towards the next generation of recurrent network models for cognitive neuroscience. *Curr Opin Neurobiol*. 2021;70:182–92. <https://doi.org/10.1016/j.conb.2021.10.015> PMID: [34844122](https://pubmed.ncbi.nlm.nih.gov/34844122/)
65. Yang GR. MultiTask Network. 2019. <https://github.com/gyyang/multitask>
66. Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat Commun*. 2018;9(1):4233. <https://doi.org/10.1038/s41467-018-06560-z> PMID: [30315158](https://pubmed.ncbi.nlm.nih.gov/30315158/)
67. Pachitariu M, Michaelos M, Stringer C. Recordings of 20,000 neurons from V1 in response to oriented stimuli. 2019. <https://doi.org/10.25378/janelia.8279387.v3>
68. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. 2009:248–55. <https://doi.org/10.1109/CVPR.2009.5206848>
69. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000;290(5500):2323–6. <https://doi.org/10.1126/science.290.5500.2323> PMID: [11125150](https://pubmed.ncbi.nlm.nih.gov/11125150/)
70. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000;290(5500):2319–23. <https://doi.org/10.1126/science.290.5500.2319> PMID: [11125149](https://pubmed.ncbi.nlm.nih.gov/11125149/)
71. Yousefi B, Khansari M, Trask R, Tallon P, Carino C, Afrasiyabi A. Density-based isometric mapping. *arXiv preprint*. 2024. <https://arxiv.org/abs/2403.02531>
72. Mienye ID, Swart TG. Deep autoencoder neural networks: a comprehensive review and new perspectives. *Archives of Computational Methods in Engineering*. 2025:1–20.
73. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9(11).
74. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*. 2019;364(6437):255. <https://doi.org/10.1126/science.aav7893> PMID: [31000656](https://pubmed.ncbi.nlm.nih.gov/31000656/)
75. Ostojic S, Fusi S. Computational role of structure in neural activity and connectivity. *Trends Cogn Sci*. 2024;28(7):677–90. <https://doi.org/10.1016/j.tics.2024.03.003> PMID: [38553340](https://pubmed.ncbi.nlm.nih.gov/38553340/)
76. Matsui T, Hashimoto T, Murakami T, Uemura M, Kikuta K, Kato T, et al. Orthogonalization of spontaneous and stimulus-driven activity by hierarchical neocortical areal network in primates. *Nat Commun*. 2024;15(1):10055. <https://doi.org/10.1038/s41467-024-54322-x> PMID: [39632809](https://pubmed.ncbi.nlm.nih.gov/39632809/)
77. Courellis HS, Minxha J, Cardenas AR, Kimmel DL, Reed CM, Valiante TA, et al. Abstract representations emerge in human hippocampal neurons during inference. *Nature*. 2024;632(8026):841–9. <https://doi.org/10.1038/s41586-024-07799-x> PMID: [39143207](https://pubmed.ncbi.nlm.nih.gov/39143207/)
78. Recanatesi S, Farrell M, Advani M, Moore T, Lajoie G, Shea-Brown E. Dimensionality compression and expansion in deep neural networks. *arXiv preprint*. 2019. <https://arxiv.org/abs/1906.00443>
79. Sharma U, Kaplan J. A neural scaling law from the dimension of the data manifold. *arXiv preprint*. 2020. <https://arxiv.org/abs/2004.10802>
80. Birdal T, Lou A, Guibas LJ, Simsekli U. Intrinsic dimension, persistent homology and generalization in neural networks. *Advances in Neural Information Processing Systems*. 2021;34:6776–89.
81. Lee JH, Jiralerspong T, Yu L, Bengio Y, Cheng E. Geometric signatures of compositionality across a language model's lifetime. *arXiv preprint*. 2024. <https://arxiv.org/abs/241001444>
82. Maheswaranathan N, Williams A, Golub M, Ganguli S, Sussillo D. Universality and individuality in neural dynamics across large populations of recurrent networks. *Adv Neural Inf Process Syst*. 2019;32.