Patterns

A scale-dependent measure of system dimensionality

Highlights

- The scale-dependent dimensionality unifies widely used measures of dimensionality

- Dynamical systems show distinct dimensionality properties at different scales

- The scale-dependent dimensionality allows us to identify critical scales of the system

- Fundamental trends in dimensionality of neural activity depend on the brain state

Authors

Stefano Recanatesi, Serena Bradde, Vijay Balasubramanian, Nicholas A. Steinmetz, Eric Shea-Brown

Correspondence

stefano.recanatesi@gmail.com (S.R.),
serena.bradde@gmail.com (S.B.),
etsb@uw.edu (E.S.-B.)

In brief

We present a scale-dependent dimensionality analysis that reveals the effective degrees of freedom in a complex system at all scales. We illustrate how, when applied to various types of data, the scale-dependent participation ratio exposes interesting geometrical patterns. Our method incorporates widely known dimensionality measures and is applicable to multivariate data in a wide range of scientific areas.

CellPress

Article

A scale-dependent measure of system dimensionality

Stefano Recanatesi,[1,7,8,*] Serena Bradde,[2,3,7,8,*] Vijay Balasubramanian,[2,6] Nicholas A. Steinmetz,[4,6] and Eric Shea-Brown[1,5,7,8,*]

[1]Center for Computational Neuroscience, University of Washington, Seattle, WA 98195, USA
[2]David Rittenhouse Laboratories, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]American Physical Society, Ridge, NY 11709, USA
[4]Department of Biological Structure, University of Washington, Seattle, WA 98195, USA
[5]Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA
[6]Senior author
[7]These authors contributed equally
[8]Lead contact
*Correspondence: stefano.recanatesi@gmail.com (S.R.), serena.bradde@gmail.com (S.B.), etsb@uw.edu (E.S.-B.)
https://doi.org/10.1016/j.patter.2022.100555

THE BIGGER PICTURE Data mining is based on the discovery of structure within data. However, such a structure is often complex. The fact that the properties of data distributions vary depending on the scale at which they are examined is a fundamental component of this complexity. For example, a manifold may appear smooth at small scales but jagged or even fractal at larger scales. This scale dependence is critical, yet it is commonly overlooked. We introduce a fundamental approach for analyzing the properties of data distributions at all scales. This single scale-dependent description enables simultaneous examination of how characteristics vary across all scales, offering insight into the structure of the data distribution. This will help us gain a better grasp of data structures and pave the way for future theoretical advances in data science.

1 2 3 4 5 **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

---

SUMMARY

A fundamental problem in science is uncovering the effective number of degrees of freedom in a complex system: its dimensionality. A system's dimensionality depends on its spatiotemporal scale. Here, we introduce a scale-dependent generalization of a classic enumeration of latent variables, the participation ratio. We demonstrate how the scale-dependent participation ratio identifies the appropriate dimension at local, intermediate, and global scales in several systems such as the Lorenz attractor, hidden Markov models, and switching linear dynamical systems. We show analytically how, at different limiting scales, the scale-dependent participation ratio relates to well-established measures of dimensionality. This measure applied in neural population recordings across multiple brain areas and brain states shows fundamental trends in the dimensionality of neural activity—for example, in behaviorally engaged versus spontaneous states. Our novel method unifies widely used measures of dimensionality and applies broadly to multivariate data across several fields of science.

INTRODUCTION

In many branches of science, complex systems are characterized by simultaneous values of many observables evolving over time. For example, the operational state of a living cell may be summarized by the expression levels of myriad pro-

teins. Likewise, the instantaneous activity levels of the many neurons in a brain region summarize its state.[1] The dynamics of these systems can be much lower dimensional. For example, at the coarsest scale, the overall dynamics of a brain area may be described just by slow fluctuations in the mean neural firing rate, i.e., a single dynamical variable.[2,3]

At an intermediate scale, the same dynamics could consist of several characteristic firing patterns evolving smoothly on a fixed d-dimensional manifold embedded in the state space. If we knew the underlying dynamical system, we could derive the relevant manifold at each scale from first principles. But in many of the most exciting complex systems becoming accessible to experimental study, our goal is to discover the dynamical system, a task that starts by determining the number of effective latent variables, i.e., the dimensionality of the system.

One approach to this problem has roots in point-set topology: a manifold is d dimensional if the number of uniformly sampled points in a region of characteristic length L scales as L^d. This fact leads to definition of the capacity dimension D_0[4] in terms of the number n of Euclidean boxes of side length ϵ needed to cover the system's trajectory in its embedding space: $D_0 = \lim_{\epsilon \to 0} \frac{\log n}{\log(1/\epsilon)}$. This intuitive quantity is difficult to compute in more than three dimensions,[5] while sampling of dynamical systems is often too coarse to directly estimate the $\epsilon \to 0$ limit.[6–8] A variation on this idea, easier to estimate in high dimensions, is the correlation dimension D_{corr},[9] determined from the scaling of the number of pairs of data points with separations less than r, in the $r \to 0$ limit. In this way, the capacity dimension and the correlation dimension both give local, fine-scale measures of dimension.[9,10]

A second class of approaches starts with the correlation matrix between observations. For example, techniques related to principal-component analysis define the effective dimension as the number of eigenmodes of the correlation matrix that capture most of the variance. Substantial literature analyzes how to choose the threshold that defines "most,"[4,11] but the resulting arbitrariness makes it challenging to define the dimension associated to different scales of observation.[12,13] A dimension can be defined more naturally from the participation ratio (PR),[1,14] which counts the effective dimensions along which data are spread as a ratio of the square of the first moment and the second moment of the eigenvalue probability density function.[15] Here, we generalize this notion to a measure of effective dimension at different observation scales and show that it interpolates between correlation dimension at small scales and PR dimension globally. We show that the new quantity has intuitive meanings when applied to dynamical systems including the Lorenz attractor and clustered systems like hidden Markov models. We then apply our method to elucidate the structure of neural population activity in different brain areas and states.

RESULTS

Generalizing the notion of dimension to all scales
Consider T observations of N observables $\boldsymbol{x}(i)$, $i = 1, \dots T$ sampled from a data distribution $\mathcal{M}$. For example, $\boldsymbol{x}(i)$ could be generated by a dynamical process $\boldsymbol{x}(t + \tau) = \mathcal{F}(\boldsymbol{x}(t))$ with specified initial conditions. The empirical covariance matrix Σ over $\boldsymbol{x}(i)$ is $\Sigma = \frac{1}{T-1}\sum_{i=1}^{T}(\boldsymbol{x}(i) - \bar{\boldsymbol{x}})(\boldsymbol{x}(i) - \bar{\boldsymbol{x}})^{\top}$, an $N \times N$ matrix with $\bar{\boldsymbol{x}} = \frac{1}{T}\sum_i^T \boldsymbol{x}(i)$. The eigenvalues of Σ are $\lambda_{i=1..N}$, and the associated spectral density is $\rho(\lambda)$. The PR dimension, D_{PR}, is defined as the ratio between the second and the first moments of the spectral density:[15]

$$D_{PR} = \frac{\text{Tr}(\Sigma)^2}{\text{Tr}(\Sigma^2)} = \frac{\left(\sum_{j=1}^{N}\lambda_j\right)^2}{\sum_{i=1}^{N}\lambda_j^2} = \frac{\left(\int \lambda\, \rho(\lambda)\, d\lambda\right)^2}{\int \lambda^2\, \rho(\lambda)\, d\lambda}. \quad \text{(Equation 2.1)}$$

D_{PR} measures the concentration of the eigenvalue distribution and quantifies how many eigenmodes are needed to substantially capture the data distribution, a similar notion to counting eigenmodes (or principal components) that capture most of the variance.

To extend D_{PR} to a scale-dependent measure of dimension, we first consider a ball of radius r, $B_i(r)$, around a point $\boldsymbol{x}(i)$. The local covariance matrix of points within this ball is $\Sigma_i(r) = \frac{1}{M-1}\sum_{\boldsymbol{x}' \in B_i(r)}(\boldsymbol{x}' - \bar{\boldsymbol{x}}')(\boldsymbol{x}' - \bar{\boldsymbol{x}}')^{\top}$, where M counts points in $B_i(r)$ and $\bar{\boldsymbol{x}}'$ is their average. In local principal-component analysis, the dominant eigenvectors of this matrix determine the local subspace in which the distribution $\mathcal{M}$ is localized. Likewise, computing the PR dimension of this covariance measures a local dimension. Averaging over all starting points yields the scale-dependent PR (sdPR)

$$D_{PR}(r) = \frac{1}{T}\sum_i \frac{(\text{Tr}\Sigma_i(r))^2}{\text{Tr}\Sigma_i(r)^2}, \quad \text{(Equation 2.2)}$$

an effective dimension up to scale r. Evaluating $D_{PR}(r)$ for the Lorenz attractor and the noisy two-dimensional (2D) spiral[4] illustrates its scale-dependent properties. For the Lorenz attractor, $D_{PR}(r)$ is roughly 2 across scales (Figure 1B), reflecting dense coverage of a 2-manifold by this chaotic system and agreeing at small scales with the Lyapunov dimension arising from its dynamics.[16] For the spiral, $D_{PR}(r)$ starts from 2 for small r, reflecting the spread of local noise, and then dips to approach ~1, reflecting the line making the spiral, and returns to 2 at large scales r, reflecting the overall embedding of the spiral (Figure 1E).

Alternatively, the correlation dimension can be generalized across scales.[17,18] Let $d_{ij} = ||\boldsymbol{x}_i - \boldsymbol{x}_j||$ be distances between samples, and define the correlation integral at distance r as $C(r) = \lim_{T \to \infty} \frac{1}{T(T-1)}\sum_{i \neq j} H(r - d_{ij}) = \int_0^r P(x)\, dx$. Here, H is the Heaviside step function, and $P(x)$ is the distribution of pairwise distances. We expect that $C(r) \propto r^d$, for small r, where d is the dimension of the manifold supporting the data $\mathcal{M}$. Then, D_{corr} is defined as $D_{corr} = \lim_{r \to 0} \frac{\log C(r)}{\log r}$. Although D_{corr} is defined as $r \to 0$, it is sometimes extended to general r as the log-log derivative $D_{corr}(r) = \frac{d \log C(r)}{d \log r} = \frac{d}{dv}\log C(e^v)$, where $v = \log r$.[4] To overcome sampling constraints as $r \to 0$, one seeks a plateau in $D_{corr}(r)$[17] at small but finite r, a valid approach if the dimension is relatively stable across a range of scales. This method effectively treats low-dimensional strange attractors,[16] but sampling remains a challenge for high-dimensional manifolds, and the problem is compounded if the effective dimension varies with scale.[6,7,19] Likewise, for any bounded dataset, $D_{corr} \to 0$ at large r simply because the data look point-like at large scales. (Technically, the correlation integral $C(r) = \int_0^r P(x) dx \to 1$ by construction at large r, so its derivative vanishes.) Indeed, applying $D_{corr}(r)$ to the Lorenz attractor, we recover the expected dimension just bigger than 2 at small

A Lorenz Attractor

B

C

D Spiral with local noise

E

F
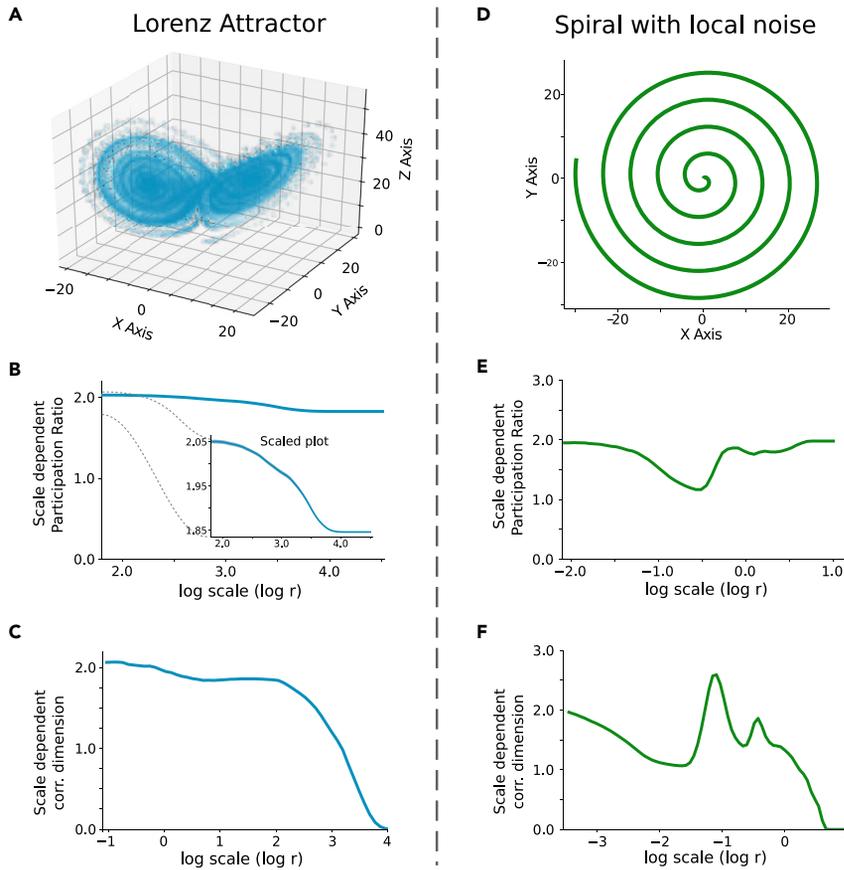
Figure 1. The scale-dependent dimensionality

(A) Lorenz attractor: 10^5 points sampled from the attractor.

(B) The scale-dependent PR (sdPR) dimension is stable across scales.

(C) The scale-dependent correlation dimension matches the Lyapunov dimension 2.05 at small scales but vanishes at large scales.

(D) Noisy spiral: 10^5 points sampled from the 2D spiral with local noise.

(E) The sdPR dimension is 2 at small scales reflecting spread due to noise, approaches 1 at intermediate scales reflecting the line, and 2 at large scales reflecting the coarse-grained spiral.

(F) The scale-dependent correlation dimension interpolates non-montonically between 2 and 0.

Details of the sdPR are in Section S1.

convolution of independent Gamma distributions for each coordinate gives the overall distribution of squared Euclidean distances between sampled points as an approximate Gamma distribution with parameters (k_s, θ_s) given by the Welch-Satterthwaite equation[20–23] simplifying here to $k_s = \frac{1}{2} \frac{(\sum_a \theta_a)^2}{\sum_a \theta_a^2}$, $\theta_s = \frac{\sum_a \theta_a^2}{\sum_a \theta_a}$. Exploiting this formula to compute the correlation integral in the $r \to 0$ limit (cf. Section S3) yields

$$D_{corr} = \frac{(\sum_a \theta_a)^2}{\sum_a \theta_a^2} = \frac{(\sum_a \lambda_a)^2}{\sum_a \lambda_a^2} = D_{PR}(r \to 0). \quad \text{(Equation 2.3)}$$

Thus, the sdPR coincides with the correlation dimension, a well-accepted local notion of dimension, at small scales. In Section S2 and Figure S1, we also relate sdPR to the Renyi dimension for small scales. We will next show in several tractable models that the sdPR—unlike the correlation and Renyi dimensions—generalizes naturally across scales.

Dimensionality across scales

We numerically computed $D_{corr}(r)$ and $D_{PR}(r)$ for isotropic multidimensional Gaussians. As shown in Figures 2A and S2A, these examples illustrate the relationship between sdPR and the correlation dimension. While D_{corr} decreases at larger scales, D_{PR} remains constant. Furthermore, we see how, limiting the sampled statistics to 50,000 points, it is not possible to achieve a plateau in D_{corr} at small scales, even for $d = 10$—this means that capturing the dimensionality of the system from the correlation dimension in Figure 2A (left) is difficult, if not impossible. In the case of 2D Gaussians with increased elongation along one of the coordinate axes (Figures 2B and S2B), the correlation dimension accurately quantifies the local two dimensions in the regime $r \to 0$ but fails to quantify the distribution's skewness, which effectively makes it 1D at large scales. In comparison, D_{PR} quantifies both the local two dimensions and the skewness, which

scales, but the value declines to zero at large scales at which the compact attractor is effectively point-like (Figure 1C). By contrast, $D_{PR}(r)$ (Figure 1B) remains roughly constant across scales, as it performs a kind of adaptive rescaling (via the denominator in Equation 2.1). Similarly, $D_{corr}(r)$ declines to zero at large scales for the noisy spiral (Figure 1F), while the PR dimension (Figure 1E) stably captures the essentially 2D large-scale structure. Note that $D_{corr}(r)$ also produces difficult-to-interpret oscillations at intermediate scales.

At small scales, we can derive a universal expression for the PR $D_{PR}(r \to 0)$ via a tangent space approximation to the data manifold that enables a new link between the PR and the correlation dimension. Approximating the local distribution as a Gaussian, $\mathcal{M} \approx \mathcal{M}^{Gauss} = (2\pi)^{-N/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^\top}$, where μ is the mean and Σ is the covariance. Rotating and centering the data so that $\mu \equiv 0$ and Σ is diagonal, each component of the sampled vectors will be normally distributed: $x^a \sim \mathcal{N}(0, \lambda_a)$, where λ_a is an eigenvalue of Σ. Thus, in this limit, $D_{PR}(r \to 0) = (\sum_a \lambda_a)^2 / \sum_a \lambda_a^2$.

In the same limit, the correlation dimension is determined by the distribution of squared Euclidean distances between sampled points. For any pair, the distance squared is a sum of squared separations in each coordinate. Since the coordinates are Gaussian distributed, the squared separations are Gamma distributed: $\Gamma(k, \theta_a)(x) = \frac{x^{k-1}e^{-x/\theta_a}}{\theta_a^k \Gamma(k)}$, with $x > 0$ where $k = 1/2$ and $\theta_a = 2\lambda_a$ for the ath coordinate direction. The

A   Multidimensional isotropic Gaussian

B   Two-dimensional skewed Gaussian

C   Multidimensional Gaussian - power law spectrum

Figure 2. The correlation dimension versus the PR dimension

(A) (Left) Scale-dependent correlation dimension and (right) sdPR of the multidimensional isotropic d-dimensional Gaussian distribution. Different lines correspond to increasing dimension d, cf. legend.

(B) (Left) Scale-dependent correlation dimension and (right) sdPR for the 2D skewed Gaussian distribution. The first eigenvalue of the diagonalized covariance matrix Σ of this distribution is $\lambda_1 = 1$, while the second varies according to the legend, determining the elongation of the distribution.

(C) (Left) Scale-dependent correlation dimension and (right) sdPR of the d-dimensional Gaussian distribution with scale-free power law spectrum. Different lines correspond to increasing dimension d, cf. legend. In this case, the dimensions have eigenvalues that are distributed according to a power law with $\alpha = 4$ so that the $k - th$ eigenvalue is $\lambda_k = k^{-\alpha}$. For each case, 50,000 points were randomly sampled. Dashed lines are extrapolations of the plotted curves in the left panel (right panel extrapolations not shown).

(C) (Right inset) PR dimension as a function of the exponent α for a scale-free spectrum of the covariance eigenvalues: $\lambda_a = \beta a^{-\alpha}$, where a is the index of the sorted eigenvalues. Then, D_{PR} at large scales converges to $D_{PR} = \frac{\zeta(\alpha)^2}{\zeta(2\alpha)}$.

Limitations of the study

The comparison between sdPR and correlation dimension highlighted advantages of sdPR over the correlation dimension. However, there also some limitations in evaluating sdPR, especially concerning its computational complexity.

The primary advantage of sdPR over correlation dimension is that it gives an estimate of the dimensionality across all scales. The disadvantages stem from the fact that $D_{PR}(r)$ is a local point-wise dimensionality estimator: $D_{PR}(r)$ is calculated as the average of multiple data points surrounding each ball of radius r (see Equation 2.2). Both the sdPR and correlation dimensions require the computation of second order statistics: the covariance matrix Σ for the sdPR and the matrix of squared distances for the correlation dimension. These two quantities have the same computational cost since they are related: $\boldsymbol{D}_{ij} = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}$. Their computational cost scales as $O(NT^2)$, where T is the number of data points and N is the dimensionality of the data. Thus, the increased computational cost of sdPR coming from the average over points is the main difference between the costs of the correlation dimension's and sdPR's.

However, the computational cost of sdPR, arising from the point-wise average in its definition, can be limited in practice, as the estimation around independent points need not be carried out over all points but just on a relatively small subset of repeats (we use $N_{repeats} = 100$). An analysis of the bias over independent repetitions yields a fast decay of the estimation bias as $1/N_{repeats}$ (cf. Section S3.4 and Dahmen et al.[25]). Given this rapid convergence, it is enough to select $1 \ll N_{repeats} \ll T$ when T is in the order of thousands or higher.

results in a reduced dimension at larger scales. The correlation dimension captures well the local two dimensions in the regime $r \to 0$ but fails to quantify the skewness of the distribution, which makes it effectively 1D at large scales. By contrast, D_{PR} quantifies both the local two dimensions and the skewness, which induces a lower dimension at larger scales.

A key further test of the sdPR is to understand how it behaves in scale-free systems. Thus, we considered multidimensional Gaussians whose covariance matrices have power-law-distributed eigenvalues, resembling structures near criticality. In these cases, sdPR equals the full rank of the covariance matrix at small scales and declines at large scales (Figures 2C and S2C). The global PR (Equation 2.1), at large scales, can be quantitatively captured; with a scale-free spectrum $\lambda_a = \beta a^{-\alpha}$, we find

$$D_{PR} = \frac{\left(\sum_i^N \lambda_i\right)^2}{\sum_i^N \lambda_i^2} = \frac{\left(\sum_a^N \beta a^{-\alpha}\right)^2}{\sum_a^N \beta a^{-2\alpha}} = \frac{\zeta(\alpha)^2}{\zeta(2\alpha)} \ , \ \text{where} \ \zeta(\alpha) \ \text{is the Riemann}$$

zeta function. This global dimension declines monotonically with the scaling exponent α: larger α creates a more skewed distribution that is effectively lower dimensional when viewed at large scales (Figure 2C, right, inset).[24] In contrast to this consistent behavior of the sdPR, the correlation dimension declines to zero at large scales, failing to capture the data dimension (Figure 2C, left). Section S3 of the supplemental experimental procedures contains another scale-free example, as well as an example based on receptive field maps of neural populations (see also Figure S3).

A
Discrete latent states (HMM)

B

C
Discrete and continuous states (SLDS)

D
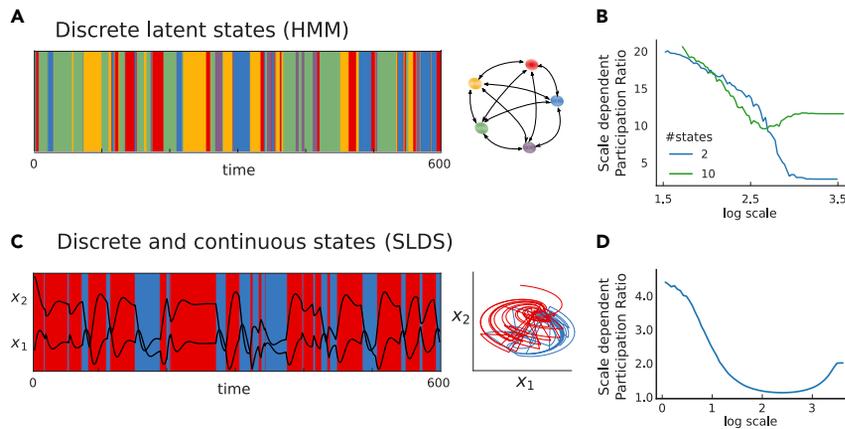
Figure 3. Scale-dependent dimensionality of hidden Markov models and switching linear dynamical systems

(A) Example of a hidden Markov model (HMM) with five hidden states. Right: latent space states diagram.

(B) sdPR for an HMM with 2 and 10 states, with observations made in a 30-dimensional observation space. At small versus large scales, the dimension is related to the structures of the observation space (where the noisy observations are high dimensional) and the state space (where the number of clusters influences the dimensionality).

(C) Example of time course of 2D latent space of switching linear dynamical system (SLDS) with two states (red and blue). Right: latent space dynamics.

(D) sdPR of the dynamics of one state in the SLDS (a rotating latent dynamical system). Observations have added Gaussian noise, and therefore the local dimensionality is higher, decays to 1 (the dimensionality of a single trajectory), and finally grows to 2, capturing the 2D circular geometry of the latent dynamics.

Figures S4 and S5 generalize the analysis of HMM and SLDS to consider different latent and observation space dimensions, noise distributions, and numbers of states.

A second consequence of sdPR arising from a local point-wise estimator is that the value of the minimum scale accessible to sdPR is higher than for the correlation dimension. The minimum distance for which the correlation dimension can be computed is $r_{min}^{corr} = \min_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{M}}(d_{ij})$. Meanwhile, $D_{PR}(r)$ computes the distance only within a ball of radius r around a subset of initial points and then averages over the initial points. Therefore, the minimum scale for sdPR is given by $r_{min}^{sdPR} = \max_{\mathbf{x}_k \in \mathcal{M}}(\min_{\mathbf{x}_i, \mathbf{x}_j \in B_k(r)}(d_{ij}))$. Thus, the correlation dimension is more granular than the sdPR dimension, $r_{min}^{corr} \leq r_{min}^{sdPR}$. Despite this apparent limitation, however, in practice, sdPR often appears to have a faster convergence to the true dimensionality at small scales, especially for higher-dimensional manifolds (cf. Figure 2A).

A measure of the local and global dimension

We next show how the sdPR can separate dimensionality driven by local dynamics that are partly influenced by noise and global dynamics that are dominated by the latent structure of a dynamical system. To illustrate, we considered data generated by an underlying hidden Markov model (HMM) with dynamics hopping between latent states (Figures 3A and S4; Section S4). The system is observed through measurements of noisy "emissions" distributed around well-separated means associated with each latent state. Locally, the dimension should be determined by the statistical spread of the data and hence equal the dimension of the emission noise in the observation space. Globally, the dimension should be related to the number of latent states and the separation of their emissions, relative to the magnitude and structure of the noise. The sdPR uncovers precisely this structure (example in Figure 3B; further details and examples in Section S4) and indicates a characteristic scale for transitioning between local and global dimensions, set by separability of the emission distributions.

For many applications, we must contend with the challenge that a "state" of the system may itself involve a characteristic dynamical trajectory with changing internal variables. Thus, we studied the sdPR of switching linear dynamical systems (SLDSs) (Figures 3C and S5; Section S4), in which the latent

states of an HMM describe dynamical trajectories following an ordinary differential equation (ODE). We found that, for intermediate scales r, $D_{PR}(r)$ identifies the system as 1D, which reflects the fact that trajectories in each state follow deterministic low-dimensional dynamics. Figure 3D demonstrates this for the case of switching rotating dynamics; the increase in dimensionality at the very smallest scales reflects the addition of emission noise. Thus, in this case, similar to Figures 1D and 1E, the local dimension is driven by stochastic observations, the global dimension is driven by the overall geometry of the dynamical system, and at intermediate scales, the structure of dynamical trajectories is revealed.

The dimensionality of neuronal data

We next applied our sdPR method to recordings of the simultaneous activity of thousands of neurons,[26] made possible by recent technological advances.[27,28] Others have examined the global dimension of the underlying systems;[1,24,29] here, we use sdPR to reveal differences in latent dynamics across scales, in different brain areas and states. Thus, we analyzed 37 Neuropixels recordings in mice from two sensory areas (visual thalamus and cortex) and two decision areas (frontal cortex and midbrain).[26,30,31] Animals were either in a "spontaneous" state (awake, no task) or in an "engaged" state (performing two alternative forced-choice tasks; Section S5).[26]

We binned neural spike counts in 100 ms windows and subdivided neurons into groups of 100 to compare across sessions sampling variable numbers of units. Intriguingly, we first found that the sdPR depended systematically on scale in the visual cortex, midbrain, and frontal cortex—decreasing at larger scales ($D_{PR}(r_{min})/D_{PR}(r_{max}) = 1.63 \pm 0.05$ mean ± SEM)—but was roughly scale invariant in the thalamus (Figure 4A). Here, distances between neural response vectors quantify scale in the functional space of neural activity, rather than physically on the cortical sheet, suggesting that activity in deeper areas is structured in groups of similar patterns spread broadly over lower-dimensional functional manifolds, like our models with skewed or scale-invariant data covariance. The thalamus, a peripheral
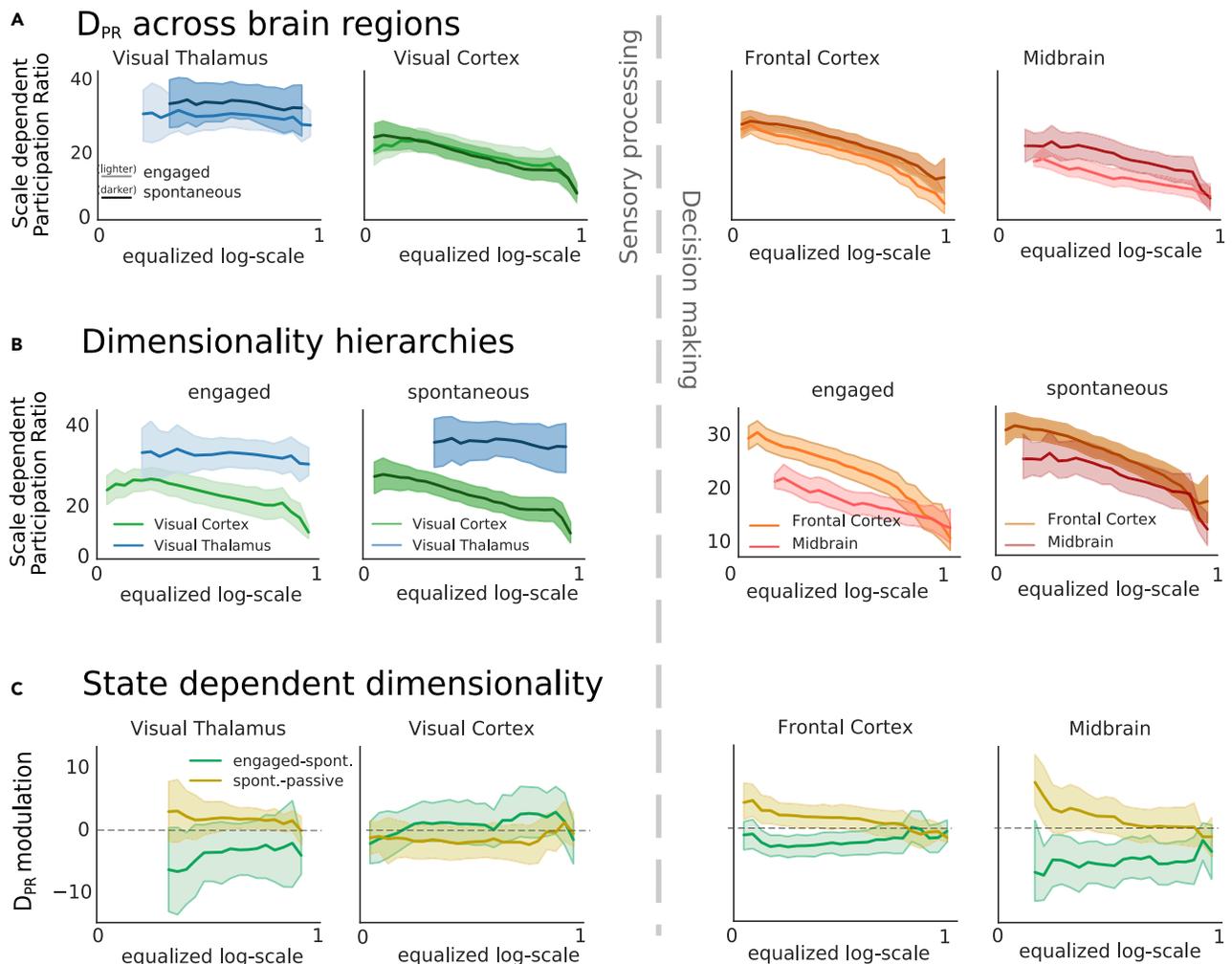
Figure 4. Scale-dependent dimensionality of neural activity
(A) sdPR across regions (scale normalized from 0 to 1; Section S5).
(B) Comparison between regions with 95% confidence intervals.
(C) Difference in dimensionality between engaged and spontaneous (green) and between spontaneous and passive conditions (yellow). Dimensionality modulation Δ^{AB} between conditions. Shading = 95% confidence interval (CI).
Extended analysis in Figures S6 and S7.

sensory area directly reflecting visual input, instead shows scale-invariant dimensionality, like our models with isotropic data covariance, and recalling the spatial scale invariance of natural images.[32] Moreover, the closeness of the thalamus to the input may explain its significantly higher dimension than the visual cortex at every scale (see Section S5 and Figures S6 and S7), despite massive expansion in the number of neurons involved in cortical as opposed to thalamic representation.[33] Interestingly, the frontal cortex, a key player in planning and executive control, had higher dimensionality than the midbrain in engaged, but not spontaneous, conditions—prominently at smaller scales accessible to our analysis despite limited data, as shown in Figure 4B.

To assess dimensionality changes within brain areas but across behavioral states, we computed differences $\Delta D_{PR}^{AB} = D_{PR}^A - D_{PR}^B$ between engaged and spontaneous conditions (Figures 4C, S6, and S7), finding significant deviation in the

frontal cortex and midbrain, which are involved in decision making, but not in sensory areas. Specifically, task-driven neural activity was lower dimensional, especially at small and intermediate scales, recalling Mazzucato et al.[34] We also considered a "passive" condition in which decision task stimuli were presented in randomized order without eliciting behaviors. The difference between spontaneous and passive was not significantly different from zero (Figure 4C, green), demonstrating that task engagement influences response dimension.

Interpreting our results on neural data (Figure 4) in light of the dynamical model results in Figure 3 leads to hypotheses about the underlying system dynamics. For example, in the engaged condition, neurons in the frontal cortex and midbrain display reduced dimensionality at intermediate scales. This suggests reduced complexity of the underlying latent dynamics (number of states or latent space dimensionality in HMM and SLDS

models) in task-driven conditions, aligning with intuitive notions of focus on a prescribed sensorimotor behavior.

DISCUSSION

Many fields employ analogs of the global PR as a measure of effective intrinsic dimension. In physics, this ratio was first introduced in atomic spectroscopy[35,36] and then used as a measure of localization in condensed matter.[15] In quantum information, a similar quantity is called "purity" and measures the degree of mixedness of states. In economics and sociology, the Herfindahl-Hirschman Index measures market concentration of an industrial sector.[37,38] In sociology, the related Simpson Index quantifies diversity,[39] while in politics, it is a measure of the effective number of parties.[40] In machine learning, the same quantity serves as a metric of expressivity for learning kernels,[41] and in neuroscience, it measures the global dimension of neural activity.[1,25,34]

Thus, the PR is used to quantify dimension in a remarkably wide variety of disciplines. However, complex systems behave differently at different scales—and thus their dimension is not necessarily characterized by a single number. Our $D_{PR}(r)$ measures a "running dimension," capturing the effective number of latent degrees of freedom required to summarize observables at different scales. Importantly, it also approaches the well-known correlation dimension at the smallest scales. Moreover, the sdPR dimension can be computed as a simple, exact functional of a system's second-order statistics, and can be derived analytically in many cases (e.g., Dahmen et al.[25] and Hu and Sompolinsky[42]). Overall, our approach may be used to analyze multivariate data in a range of domains, and we expect that it will reveal new geometrical aspects of data across many fields.

EXPERIMENTAL PROCEDURES

Resource availability
Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Stefano Recanatesi (stefano.recanatesi@gmail.com).
Materials availability
This study did not generate new unique reagents or materials.
Data and code availability
Original code has been deposited at https://github.com/ubcbraincircuits/accelnet-jupyternotebooks-dimensionality and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request. This paper analyzes existing, publicly available electrophysiology data deposited in Neurodata Without Borders (NWB) format at https://figshare.com/articles/steinmetz/9598406.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100555.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

S.R. and S.B. conducted the analysis; N.A.S. performed the experiments; S.R., S.B., V.B., N.A.S., and E.S.-B. designed the research and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

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

2. Harris, K.D., and Thiele, A. (2011). Cortical state and attention. Nat. Rev. Neurosci. *12*, 509–523. https://doi.org/10.1038/nrn3084.

3. Okun, M., Steinmetz, N.A., Cossell, L., Iacaruso, M.F., Ko, H., Barthó, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., and Harris, K.D. (2015). Diverse coupling of neurons to populations in sensory cortex. Nature *521*, 511–515. https://doi.org/10.1038/nature14273.

4. Lee, J.A., and Verleysen, M. (2007). Nonlinear Dimensionality Reduction, 1st edition (Springer Publishing Company, Incorporated).

5. Eckmann, J.-P., and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. In The Theory of Chaotic Attractors (Springer), pp. 273–312.

6. Ding, M., Grebogi, C., Ott, E., Sauer, T., and Yorke, J.A. (1993). Plateau onset for correlation dimension: when does it occur? Phys. Rev. Lett. *70*, 3872–3875. https://doi.org/10.1103/physrevlett.70.3872.

7. Michalak, K.P. (2014). How to estimate the correlation dimension of high-dimensional signals? Chaos: An Interdisciplinary Journal of Nonlinear Science *24*, 033118. https://doi.org/10.1063/1.4891185.

8. Michalak, K.P. (2018). Estimating correlation dimension of high-dimensional signals - quick algorithm. AIP Adv. *8*, 105201. https://doi.org/10.1063/1.5013255.

9. Grassberger, P., and Procaccia, I. (1983). Characterization of strange attractors. Phys. Rev. Lett. *50*, 346–349. https://doi.org/10.1103/physrevlett.50.346.

10. Facco, E., d'Errico, M., Rodriguez, A., and Laio, A. (2017). Estimating the intrinsic dimension of datasets by a minimal neighborhood information. Sci. Rep. *7*, 12140. https://doi.org/10.1038/s41598-017-11873-y.

11. Camastra, F., and Staiano, A. (2016). Intrinsic dimension estimation: advances and open problems. Inf. Sci. *328*, 26–41. https://doi.org/10.1016/j.ins.2015.08.029.

12. Fukunaga, K., and Olsen, D.R. (1971). An algorithm for finding intrinsic dimensionality of data. IEEE Trans. Comput. *20*, 176–183. https://doi.org/10.1109/t-c.1971.223208.

13. Kambhatla, N., and Leen, T.K. (1997). Dimension reduction by local principal component analysis. Neural Comput. *9*, 1493–1516. https://doi.org/10.1162/neco.1997.9.7.1493.

14. Rosenbluth, G. (1955). Measures of concentration. In Business Concentration and Price Policy (Princeton University Press), pp. 57–99.

15. Kramer, B., and MacKinnon, A. (1993). Localization: theory and experiment. Rep. Prog. Phys. *56*, 1469–1564. https://doi.org/10.1088/0034-4885/56/12/001.

16. Kuznetsov, N.V., Mokaev, T.N., Kuznetsova, O.A., and Kudryashova, E.V. (2020). The lorenz system: hidden boundary of practical stability and the lyapunov dimension. Nonlinear Dynam. *102*, 713–732. https://doi.org/10.1007/s11071-020-05856-4.

17. Albano, A.M., Muench, J., Schwartz, C., Mees, A.I., and Rapp, P.E. (1988). Singular-value decomposition and the grassberger-procaccia algorithm. Phys. Rev. A *38*, 3017–3026.

18. Lacasa, L., and Gómez-Gardeñes, J. (2013). Correlation dimension of complex networks. Phys. Rev. Lett. *110*, 168703. https://doi.org/10.1103/physrevlett.110.168703.

19. Guerrero, A., and Smith, L.A. (2003). Towards coherent estimation of correlation dimension. Phys. Lett. *318*, 373–379. https://doi.org/10.1016/j.physleta.2003.09.023.

20. Ballico, M. (2000). Limitations of the welch-satterthwaite approximation for measurement uncertainty calculations. Metrologia *37*, 61.

21. Huang, H. (2016). On the welch-satterthwaite formula for uncertainty estimation: a paradox and its resolution. Cal Lab the International Journal of Metrology *23*, 20–28.

22. Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics Bull. *2*, 110. https://doi.org/10.2307/3002019.

23. Welch, B.L. (1947). The generalization of 'Student's' problem when several different population variances are involved. Biometrika *34*, 28. https://doi.org/10.2307/2332510.

24. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional geometry of population responses in visual cortex. Nature *571*, 361–365. https://doi.org/10.1038/s41586-019-1346-5.

25. Dahmen, D., Recanatesi, S., Ocker, G.K., Jia, X., Helias, M., and Shea-Brown, E. (2020). Strong coupling and local control of dimensionality across brain areas. Preprint at bioRxiv. https://doi.org/10.1101/2020.11.02.365072.

26. Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273. https://doi.org/10.1038/s41586-019-1787-x.

27. Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236. https://doi.org/10.1038/nature24636.

28. Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2020). Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings. Preprint at bioRxiv. https://doi.org/10.1101/2020.10.27.358291.

29. Litwin-Kumar, A., Harris, K.D., Axel, R., Sompolinsky, H., and Abbott, L. (2017). Optimal degrees of synaptic connectivity. Neuron *93*, 1153–1164.e7. https://doi.org/10.1016/j.neuron.2017.01.030.

30. Felsen, G., and Mainen, Z.F. (2012). Midbrain contributions to sensorimotor decision making. J. Neurophysiol. *108*, 135–147. https://doi.org/10.1152/jn.01181.2011.

31. Horwitz, G.D., Batista, A.P., and Newsome, W.T. (2004). Representation of an abstract perceptual decision in macaque superior colliculus. J. Neurophysiol. *91*, 2281–2296. https://doi.org/10.1152/jn.00872.2003.

32. Schwartz, O., and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. Nat. Neurosci. *4*, 819–825. https://doi.org/10.1038/90526.

33. Olshausen, B.A. (2003). Principles of image representation in visual cortex. Visual Neurosci. *2*, 1603–1615.

34. Mazzucato, L., Fontanini, A., and La Camera, G. (2016). Stimuli reduce the dimensionality of cortical activity. Front. Syst. Neurosci. *10*, 11. https://doi.org/10.3389/fnsys.2016.00011.

35. Bell, R.J., and Dean, P. (1970). Atomic vibrations in vitreous silica. Discuss. Faraday Soc. *50*, 55. https://doi.org/10.1039/df9705000055.

36. Last, B.J., and Thouless, D.J. (1974). Evidence for power law localization in disordered systems. J. Phys. C Solid State Phys. *7*, 699–715. https://doi.org/10.1088/0022-3719/7/4/008.

37. Hall, M., and Tideman, N. (1967). Measures of concentration. J. Am. Stat. Assoc. *62*, 162–168. https://doi.org/10.1080/01621459.1967.10482897.

38. Hirschman, A.O. (1964). The paternity of an index. Am. Econ. Rev. *54*, 761–762.

39. Simpson, E.H. (1949). Measurement of diversity. Nature *163*, 688. https://doi.org/10.1038/163688a0.

40. Laakso, M., and Taagepera, R. (1979). effective" number of parties: a measure with application to west europe. Comp. Polit. Stud. *12*, 3–27. https://doi.org/10.1177/001041407901200101.

41. Donini, M., and Aiolli, F. (2017). Learning deep kernels in the space of dot product polynomials. Mach. Learn. *106*, 1245–1269. https://doi.org/10.1007/s10994-016-5590-8.

42. Hu, Y., and Sompolinsky, H. (2020). The spectrum of covariance matrices of randomly connected recurrent neuronal networks. Preprint at bioRxiv. https://doi.org/10.1101/2020.08.31.274936.