



Published in final edited form as:

*Nat Neurosci.* 2014 November ; 17(11): 1500–1509. doi:10.1038/nn.3776.

## Dimensionality reduction for large-scale neural recordings

John P Cunningham<sup>1</sup> and Byron M Yu<sup>2,3,4</sup>

<sup>1</sup>Department of Statistics, Columbia University, New York, New York, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Department of Biomedical Engineering, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

<sup>3</sup>Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

<sup>4</sup>Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

### Abstract

Most sensory, cognitive and motor functions depend on the interactions of many neurons. In recent years, there has been rapid development and increasing use of technologies for recording from large numbers of neurons, either sequentially or simultaneously. A key question is what scientific insight can be gained by studying a population of recorded neurons beyond studying each neuron individually. Here, we examine three important motivations for population studies: single-trial hypotheses requiring statistical power, hypotheses of population response structure and exploratory analyses of large data sets. Many recent studies have adopted dimensionality reduction to analyze these populations and to find features that are not apparent at the level of individual neurons. We describe the dimensionality reduction methods commonly applied to population activity and offer practical advice about selecting methods and interpreting their outputs. This review is intended for experimental and computational researchers who seek to understand the role dimensionality reduction has had and can have in systems neuroscience, and who seek to apply these methods to their own data.

---

A central tenet of neuroscience is that the remarkable computational abilities of our brains arise as a result of populations of interconnected neurons. Indeed, we find ourselves at an exciting moment in the history of neuroscience, as the field is experiencing rapid growth in the quantity and complexity of the recorded neural activity. Many groups have begun to adopt multi-electrode<sup>1</sup> and optical<sup>2</sup> recording technologies that can monitor the activity of many neurons simultaneously in cortex and, in some cases, in deeper structures. Ongoing development of recording technologies promises to increase the number of simultaneously recorded neurons by orders of magnitude<sup>3</sup>. At the same time, massive increases in

---

© 2014 Nature America, Inc. All rights reserved

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to J.P.C. (jpc2181@columbia.edu) or B.M.Y. (byronyu@cmu.edu).

**COMPETING FINANCIAL INTERESTS** The authors declare no competing financial interests.

computational power and algorithmic development have enabled advanced multivariate analyses of neural population activity, where the neurons may be recorded either sequentially or simultaneously.

These technological advances have enabled researchers to reconsider the types of scientific questions that are being posed and how neural activity is analyzed, even with classical behavioral tasks and in brain areas that have been studied for decades. Indeed, many studies of neural systems are undergoing a paradigm shift from single-neuron to population-level hypotheses and analyses. We begin this review by discussing three scientific motivations for considering a neural population jointly, rather than on a single-neuron basis: single-trial hypotheses requiring statistical power, hypotheses of population response structure and exploratory analyses of large data sets. Critically, we show that there are settings in which data fundamentally cannot be understood on a single-neuron basis, whether as a result of neural spiking variability or a hypothesis about neural mechanism that depends on how the responses of multiple neurons covary.

The object of this review is to focus on one class of statistical methods, dimensionality reduction, which is well-suited for analyzing neural population activity. Dimensionality reduction methods produce low-dimensional representations of high-dimensional data, where the representation is chosen to preserve or highlight some feature of interest in the data. These methods have begun to reveal tantalizing evidence of the neural mechanisms underlying various phenomena, including the selection and integration of sensory input during decision-making in prefrontal cortex<sup>4</sup>, the ability of premotor cortex to prepare movements without executing them<sup>5</sup>, and odor discrimination in the olfactory system<sup>6</sup>. Dimensionality reduction has also been fruitfully applied to population recordings in other studies of decision-making<sup>7-9</sup>, the motor system<sup>10-12</sup> and the olfactory system<sup>13,14</sup>, as well as in working memory<sup>15,16</sup>, visual attention<sup>17</sup>, the auditory system<sup>18</sup>, rule learning<sup>19</sup>, speech<sup>20</sup> and more. We introduce dimensionality reduction and bring together previous studies that have used these methods to address each of the three scientific motivations for population analyses. Because the use of dimensionality reduction is still relatively new in systems neuroscience, we then present methodological details and practical considerations.

Much of this work in neuroscience has developed in the last decade: as presciently noted by Brown *et al.*<sup>21</sup>, “the future challenge is to design methods that truly allow neuroscientists to perform multivariate analyses of multiple spike train data”. Dimensionality reduction is one important way in which many researchers have answered and will continue to answer this challenge.

## Scientific motivation of population analyses

The growth in scale and resolution of recording technologies brings with it challenges for the analysis of neural activity. Consider a didactic example of action potentials from multiple simulated neurons ‘recorded’ across multiple experimental trials of multiple experimental conditions (Fig. 1a). As the number of neurons, trials and conditions grows, it becomes increasingly challenging to extract meaningful structure from these spike trains. Indeed, the classic approach of averaging spike trains across putatively similar trials and

smoothing across time (Fig. 1b) may still yield responses that are difficult to interpret. Despite this apparent complexity, there is also great scientific opportunity in studying a population of neurons together; here we discuss these scientific motivations.

### Single-trial statistical power

If the neural activity is not a direct function of externally measurable or controllable variables (for example, if activity is more a reflection of internal processing than stimulus drive or measurable behavior), the time course of neural responses may differ substantially on nominally identical trials. One suspects this to be especially true of cognitively demanding tasks that involve attention, decision-making and more. In this setting, averaging responses across trials may obscure the neural time course of interest, and single-trial analyses are therefore essential. By recording the response of a single neuron, it is usually difficult to identify the moment-by-moment fluctuations of these types of internal cognitive processes. However, if multiple neurons are recorded simultaneously, one can leverage statistical power across neurons to extract a succinct summary of the population activity on individual experimental trials<sup>12,22–24</sup>.

For example, consider a decision-making task in which the subject might abruptly change his/her mind or vacillate between possible choices on individual trials<sup>25,26</sup>. If these switches occur at different times on different trials, the act of trial averaging will obscure the switching times of each trial. At worst, trial averaging can mislead scientific interpretation: in this example, abrupt, but temporally variable, switches appear as a slow transition when averaged, suggestive of a different neural mechanism. Population recordings critically address this shortcoming of trial averaging: one can consider multiple neurons on a single trial, rather than a single neuron on multiple trials, to gain the statistical power necessary to extract de-noised single-trial neural time courses. These time courses can then be related to the subject's behavior on a trial-by-trial basis, potentially leading to new insights about the neural basis of decision-making<sup>7,9,24,26</sup>. Below, we will show that dimensionality reduction methods are a natural choice for this statistical analysis, as used in the above studies.

### Population response structure

Population analyses are necessary in settings in which there may be neural mechanisms that involve coordination of responses across neurons. These mechanisms exist only at the level of the population and not at the level of single neurons, such that single-neuron responses can appear hopelessly confusing or, worse, can mislead the search for the true biological mechanism<sup>27</sup>. Indeed, recordings in higher level brain areas<sup>16</sup>, as well as areas closer to sensory inputs<sup>28</sup> and motor outputs<sup>29</sup>, have yielded highly heterogeneous and complex single-neuron responses, both across neurons and across experimental conditions. In some cases, single-neuron responses may bear no obvious moment-by-moment relationship with the sensory input or motor output that can be externally measured. Classically, such heterogeneity has been considered to be a result of biological noise or other confounds, and often researchers study only neurons that 'make sense' in terms of externally measurable quantities. However, this single-neuron complexity may be the realization of a coherent and testable neural mechanism that exists only at the level of the population.

To make this concept concrete, consider a hypothetical neural circuit as a dynamical system, in the sense that the activity of its constituent neurons changes over time<sup>30</sup>. Just as the motion of a bouncing ball is governed by Newton's laws, the activity in this neural circuit is governed by dynamical rules (for example, point attractors, line attractors or oscillations). Although single-neuron responses certainly express aspects of these dynamical rules, the critical conceptual point is that neither the single-neuron responses in isolation nor the true mechanism can be understood without the population of neurons. By analyzing the recorded neurons together, one can test for the presence or absence of a neural mechanism, be that a dynamical process or some other type of population activity structure. Dimensionality reduction is a key statistical tool for forming and evaluating hypotheses about population activity structure<sup>4-6,11,16</sup>.

### Exploratory data analysis

Studying a population of neurons together facilitates data-driven hypothesis generation. There is a subtle, but important, difference between this and the previous motivation. Population response structure is concerned with mechanisms that exist at the population level and are not interpretable as single-neuron hypotheses or mechanisms. On the other hand, exploratory data analysis involves visualizing a large amount of data, which can help to generate hypotheses regarding either single neurons or the population. When the neurons show heterogeneous response properties, it can be challenging to interpret all responses simultaneously and cohesively (Fig. 1). Consideration of the population as a whole provides a way in which all of the data (across neurons, conditions, trials and time) can be interpreted together. This step provides an initial assessment of the salient features of the data and can guide subsequent analyses. Furthermore, visualization is an efficient way to perform sanity checks on large data sets (for example, to see that neural activity is more similar across trials of the same experimental condition than across trials of different conditions, or to check for recording stability across an experimental session), which facilitates rapid iteration of the experimental design. Dimensionality reduction, by giving a low-dimensional summary of the high-dimensional population activity, is a natural approach for performing exploratory data analysis.

### Intuition behind dimensionality reduction

Dimensionality reduction is typically applied in settings in which there are  $D$  measured variables, but one suspects that these variables covary according to a smaller number of explanatory variables  $K$  (where  $K < D$ ). Dimensionality reduction methods discover and extract these  $K$  explanatory variables from the high-dimensional data according to an objective that is specific to each method. These explanatory variables are often termed latent variables because they are not directly observed. Typically, any data variance not captured by the latent variables is considered to be noise. In this light, dimensionality reduction is like many statistical methods: it provides a parsimonious description of statistical features of interest and discards some aspects of the data as noise.

In the case of neural population activity,  $D$  usually corresponds to the number of recorded neurons. Because the recorded neurons belong to a common underlying network, the responses of the recorded neurons are likely not independent of each other. Thus, fewer

latent variables may be needed to explain the population activity than the number of recorded neurons. The latent variables can be thought of as common input or, more generally, as the collective role of unobserved neurons in the same network as the recorded neurons. Furthermore, many dimensionality reduction methods take the view that the time series of action potentials emitted by a single neuron can be represented by an underlying, time-varying firing rate, from which the action potentials are generated in a stochastic manner<sup>31–33</sup>. This is a prevalent view in neuroscience, whether it is stated implicitly (for example, averaging spike trains across trials to estimate a time-varying firing rate) or explicitly (for example, statistical models of spike trains<sup>34</sup>). Previous studies have suggested that the stochastic component tends to be Poisson-like<sup>35</sup>, and we refer to it as spiking variability. The goal of dimensionality reduction is to characterize how the firing rates of different neurons covary while discarding the spiking variability as noise. Thus, each of the  $D$  neurons can be thought of as providing a different, noisy view of an underlying, shared neural process, as captured by the  $K$  latent variables. The latent variables define a  $K$ -dimensional space that represents shared activity patterns that are prominent in the population response.

To illustrate, consider the case of  $D = 3$  neurons. We first define a high-dimensional space in which each axis represents the firing rate of a neuron ( $r_1$ ,  $r_2$ , and  $r_3$ ; Fig. 2). In this framework, each vector  $[r_1, r_2, r_3]$  of population activity corresponds to a point in this space. We can then ask what low-dimensional ( $K < D$ ) space explains these data well. In Figure 2, the points lie on a plane (shaded gray,  $K = 2$ ) and trace out a trajectory over time. Each time point  $t$  corresponds to a single point in the high-dimensional firing rate space  $[r_1(t), r_2(t), r_3(t)]$ . Note that time is not plotted on any of the axes; each axis represents the firing rate of one neuron, and time evolves implicitly across the trajectory.

There are two complementary ways to think about the relationship between the latent variables and the population activity for linear dimensionality reduction. First, the population activity can be reconstructed by taking a weighted combination of the latent variables, where the weights are determined by the dimensionality reduction method (Fig. 2). Each latent variable time course ( $s_1(t)$  and  $s_2(t)$ ) can be thought of as a temporal basis function: a characteristic pattern of covarying activity shared by different neurons. The first and second columns of the weight matrix specify the  $s_1$  and  $s_2$  axes in the three-dimensional space, respectively (Fig. 2). Thus, one can think of each weight as specifying how much of each temporal basis pattern to use when reconstructing the response of each neuron. Second, the latent variables can be considered low-dimensional readouts or descriptions of the high-dimensional population activity (in  $[s_1, s_2]$  space; Fig. 2). Each latent variable can be obtained simply by taking a weighted combination of the activity of different neurons.

In general, the points in  $[r_1, r_2, r_3]$  space will not lie exactly in a plane. In this case, dimensionality reduction attempts to find latent variables that can reconstruct the population activity as well as possible. The reconstructed activity can be interpreted as the de-noised firing rate for each neuron.

## Scientific studies using dimensionality reduction

Having established the intuition of dimensionality reduction, we discuss the uses of these methods in the neuroscience literature and the insights they have revealed, organized by the three scientific motivations for analyzing neural populations.

### Single-trial statistical power

A growing body of work has leveraged statistical power across multiple neurons to characterize the population activity on individual experimental trials. A prominent example is the study of visual attention, which likely varies from moment to moment despite the best efforts of the experimenters. To study the neural mechanisms underlying visual attention, one group<sup>17</sup> recorded from a population of neurons in monkey area V4 during a change-detection task. They then established a single-trial measure of attention by projecting the population activity onto a one-dimensional 'attention axis' (Fig. 3a), which is dimensionality reduction onto a line ( $K = 1$ ). The attention axis was defined by the mean response of each neuron in each attention condition. Notably, the authors found that the projection onto the attention axis predicted behavioral performance on a trial-by-trial basis (Fig. 3a), an effect that was not possible to see at the level of individual neurons. The key to this finding was the projection, which leveraged statistical power across the entire recorded population to estimate an underlying attentional state on a single-trial basis.

There are several other key experimental contexts in which dimensionality reduction has been applied to population activity to reveal single-trial neural phenomena, including studies of decision-making<sup>9,24,26</sup>, rule learning<sup>19</sup>, motor planning<sup>36</sup> and stimulus localization<sup>22</sup>. In these studies, dimensionality reduction was performed using the population responses alone, without referencing the subject's behavior. The behavior was then used to validate the extracted latent variables on a single-trial basis. A similar approach can be used to study how population activity differs on trials with aberrant behavior (that is, error trials)<sup>7,9,10</sup>.

Dimensionality reduction is also valuable for studying spontaneous activity, where no notion of a repeatable trial exists and must therefore be analyzed on a single-trial basis. By definition, spontaneous activity involves fluctuations of the population activity that are not directly controlled by the experimenter. To characterize spontaneous activity, dimensionality reduction can be applied to extract a low-dimensional network state at each moment in time<sup>18,37-39</sup>. This facilitates the comparison of spontaneous activity to population activity during sensation<sup>10,18</sup> and action<sup>10,37</sup>.

### Population response structure

Another key use of dimensionality reduction is to test scientific hypotheses that are sensible only at the level of the population. Population structure hypotheses have been actively pursued in several different systems, including the prefrontal cortex<sup>4,8,15,16,40</sup>, the motor system<sup>5,11,41-43</sup> and the olfactory system<sup>6,13,14,28,44-48</sup>. A common theme across these systems is that, although neural responses may appear hopelessly complex at the level of individual neurons, simpler organizing principles exist at the level of the population.

In the prefrontal cortex, one study<sup>4</sup> examined the representation of both relevant and irrelevant stimulus information in a decision-making task, which comprised 72 different experimental conditions involving the motion and color of dots on a screen (Fig. 3b). During this task, the authors recorded over 1,000 neurons in the prefrontal cortex, yielding a large database of responses that could not be easily summarized or understood, a prototypical example of single-neuron complexity. Given the lack of simple interpretation at the level of individual neurons, they asked whether the confounding single-neuron responses could be understood as views of a simple dynamical process at the population level. A dimensionality reduction method was designed to identify shared latent variables, each of which explained an external covariate (the subject's choice, dot motion, dot color or task context). Notably, the low-dimensional representation was a simple projection of the population activity, and not a decoded estimate of these covariates. Applying this method to experimental data, the authors found that the population activity was consistent with a low-dimensional dynamical process involving a line attractor (Fig. 3b), which could not have been identified by examining single neurons in isolation. Furthermore, they found that the population activity surprisingly carried both relevant and irrelevant stimulus information, and that 'gating' could be implicitly achieved by a readout mechanism along task-dependent directions in the population space, an inherently population-level concept. Their use of dimensionality reduction also provided a bridge between population recordings and network theory<sup>49</sup>, in which both the experimental data and the model indicated a similar dynamical process for information gating. The mechanism also predicted that the dynamical process would have a different orientation in the population space for different task contexts, which can be tested experimentally to validate or invalidate this mechanism. This connection between experiment and theory was enabled by the judicious use of dimensionality reduction to examine population response structure.

In the motor system, another group<sup>11</sup> studied reach preparation and execution by recording hundreds of neurons in the primary motor cortex during a task with 108 different experimental conditions (Fig. 3c). As in the previous example<sup>4</sup>, the heterogeneity of the single-neuron responses was difficult to interpret. By applying dimensionality reduction, they found a coherent mechanism at the level of the population: preparatory activity sets the initial state of a dynamical process, which unfolds as the movement is executed. This dynamical structure cannot be understood with single-neuron responses alone, and population analysis was required to reveal this lawful coordination between two phases of the task (Fig. 3c). This structure further predicted a lack of correlation in neuronal tuning between the preparatory and execution-related phases of the task at the level of single neurons, as had been previously observed<sup>29</sup>. More recently, dimensionality reduction was used to show that population-level mechanisms similar to the readout directions for relevant information underlie the implicit gating of movement in the motor system during preparatory periods<sup>5</sup>.

Finally, the responses of neurons in the olfactory system to different odors have also been considered through the lens of population response structure<sup>28</sup>. Here, diverse time courses across neurons and odor stimuli have confounded attempts to understand the basic encoding of these odors. This task becomes increasingly difficult as the number of neurons and stimulus conditions increases. Dimensionality reduction has been applied to the activity of

approximately 100 of the 800 neurons in the locust antennal lobe, and the population activity traces out loops that are organized by stimulus condition. In particular, the orientation of the loop is related to the odor identity, and the size of the loop is related to the odor concentration<sup>44</sup>. These results have been extended in important ways to elucidate the temporal dynamics of this encoding<sup>6</sup>, encoding of upstream olfactory receptor neurons<sup>48</sup>, temporally structured odor stimuli<sup>45</sup>, rapid temporal fluctuations of odors<sup>47</sup>, coding of overlapping odors<sup>13</sup>, coding of mixtures of odors<sup>14</sup> and more.

### Exploratory data analysis

Dimensionality reduction is also a useful tool for exploratory data analysis on large neural data sets. A telling example comes from optical recordings throughout the entire brain of a larval zebrafish at cellular resolution during motor adaptation<sup>50</sup>. The quantity of recorded data in this setup and similar<sup>3</sup> is staggering: upwards of one terabyte of recorded activity per hour from over 80,000 neurons<sup>51</sup>. To visualize and begin to understand this data, the authors applied dimensionality reduction to reveal four distinct types of neural dynamics (Fig. 3d). They then connected this insight back to the neural architecture and found that each of these dynamical regimes corresponded to single neurons in distinct brain areas, suggesting a new role for each of these regions. Thus, dimensionality reduction allowed the formation of a new hypothesis about the response properties of single neurons.

Even when recording from a more modest number of neurons, there is a need for methods to visualize the population activity in a concise fashion. Dimensionality reduction has been used for data visualization and hypothesis generation in various brain areas, including the motor cortex<sup>36,42,52,53</sup>, hippocampus<sup>54</sup>, frontal cortex<sup>55</sup>, auditory cortex<sup>56</sup>, prefrontal cortex<sup>57</sup>, striatal cortex<sup>58</sup> and the olfactory system<sup>59</sup>. Although most of these studies proceeded to test hypotheses using the raw neural activity (without dimensionality reduction), the use of dimensionality reduction was vital for generating the hypotheses in the first place and guiding subsequent analyses<sup>60</sup>. The interplay between hypothesis generation and data analysis, as facilitated by dimensionality reduction, may become increasingly essential as the size of neural data sets grows.

### Selecting a dimensionality reduction method

As the previous section illustrates, there are many dimensionality reduction methods, each differing in the statistical structure it preserves and discards. Although many methods have deep similarities<sup>61</sup>, as with any statistical technique, the choice of method can have significant bearing on the scientific interpretations that can be made. For this reason, and because the use of dimensionality reduction is still relatively new in systems neuroscience, the following two sections are presented to introduce new users to these methods, to help existing users with method choice and interpretation, and to describe potential pitfalls of each choice. We describe the dimensionality reduction methods that are most commonly applied to neural activity (Table 1) and provide guidance for their appropriate use. Although the descriptions below focus on electrical recordings of spike trains, these methods can be applied equally well to fluorescence measurements from optical imaging<sup>7,9,50</sup> and other types of neural signals.



## Basic covariance methods

Principal component analysis (PCA) and factor analysis (FA) are two of the most basic and well-used dimensionality reduction methods. For illustrative purposes, consider the case of  $D = 2$  neurons and  $K = 1$  latent variables (Fig. 4a). We begin by forming high-dimensional vectors of raw or processed (for example, trial averaged) spike counts. Each data vector corresponds to a dot in Figure 4a. PCA identifies an ordered set of orthogonal directions that captures the greatest variance in the data. The direction of greatest variance is denoted by  $s_1$ . The orthogonal  $s_2$  axis (not shown) is the direction that captures the least variance. The data can then be projected onto the  $s_1$  axis, forming a one-dimensional data set that best preserves the data covariance (Fig. 4a).

Although capturing the largest amount of variance may be desirable in some scenarios, a caveat is that the low-dimensional space identified by PCA captures variance of all types, including firing rate variability and spiking variability. Because spiking variability can obscure the interpretation of the latent variables, PCA is usually applied to trial-averaged (and, in some cases, temporally smoothed) spike counts, where the averaging removes much of the spiking variability in advance. If one seeks to analyze raw spike counts, FA can be used to better separate changes in firing rate from spiking variability. FA identifies a low-dimensional space that preserves variance that is shared across neurons (considered to be firing rate variability) while discarding variance that is independent to each neuron (considered to be spiking variability)<sup>10</sup>. FA can be seen as PCA with the addition of an explicit noise model that allows FA to discard the independent variance for each neuron.

## Time series methods

If the data form a time series, one can leverage the sequential nature of the data to provide further de-noising and to characterize the temporal dynamics of the population activity. Although there have been many important developments of time series methods tailored for multi-neuronal spike trains (see refs. 62–67 for examples), we focus on a subset of these methods that identify low-dimensional structure in an unsupervised fashion (that is, where some or all of the predictors of neural activity are not directly observed).

There are several dimensionality reduction methods available for time series: hidden Markov models (HMM)<sup>22–24,26,68–70</sup>, kernel smoothing followed by a static dimensionality reduction method, Gaussian process factor analysis (GPFA)<sup>12</sup>, latent linear dynamical systems (LDS)<sup>71–76</sup> and latent nonlinear dynamical systems (NLDS)<sup>77,78</sup>. All of these methods return low-dimensional, latent neural trajectories that capture the shared variability across neurons for each high-dimensional time series. An HMM is applied in settings where the population activity is believed to jump between discrete states, whereas all of the other methods identify smooth changes in firing rates over time (where the degree of smoothness is determined by the data). A common way to characterize trial-averaged responses across a population of neurons is to average across trials and temporally smooth each neuron's response, and to then apply PCA. This yields a neural trajectory for each experimental condition and facilitates the comparison of population activity across conditions<sup>4,6,11</sup>. In contrast, HMM, GPFA, LDS and NLDS are typically applied to single-trial population activity. This yields single-trial neural trajectories, which facilitate the comparison of

population activity across trials<sup>10</sup>, and a low-dimensional dynamics model, which characterizes how the population activity evolves over time. These methods are particularly appropriate for single-trial population activity because they have explicit noise models (akin to FA).

As a cautionary note, when interpreting the neural trajectories, it is important to understand the steps and assumptions involved in extracting them from the high-dimensional population activity. For methods with an explicit dynamics model, its parameters are first fit to a set of (training) trials. Then, a low-dimensional trajectory can be extracted by making a statistical tradeoff between the dynamics model and the noisy (test) data. Thus, a particular low-dimensional trajectory may be as much a reflection of the dynamics model as of the data. For example, the dynamics model in GPFA is stationary and encourages the trajectories to be smooth, whereas that in LDS and NLDS is generally nonstationary and encourages the trajectories to follow particular dynamical motifs. For this reason, we recommend a simple first approach such as PCA on smoothed, trial-averaged data or GPFA on single-trial data, which can then guide the choice of a directed dynamics model, such as LDS or NLDS. In all cases, the extracted trajectories should be interpreted with caution in the context of the type of structure encouraged by the dynamics model.

If one seeks trajectories that are a projection of the data (that does not require a statistical tradeoff with a dynamics model), one can use an orthogonal projection (akin to PCA) to extract low-dimensional trajectories after having identified the low-dimensional space using a method that involves a dynamics model. The extracted trajectories are then simply a projection of the data and have not been constrained by a dynamics model, with the tradeoff that one gives up de-noising of the trajectories that would be provided by the dynamics model. Such a method was developed to investigate the rotational structure of neural population dynamics<sup>11</sup>.

### Methods with dependent variables

In many experimental settings, each data point in the high-dimensional firing rate space has an associated label of one or more dependent variables. These dependent variables may correspond to experimental parameters (for example, stimulus identity), the subject's behavior (for example, decision identity) or a time index. A possible objective of dimensionality reduction is to project the data such that differences in these dependent variables are preserved, in contrast with all of the methods described above that discover structure in the population activity in an unsupervised fashion. If each data point belongs to one of  $G$  groups (for example, experimental conditions), then linear discriminant analysis (LDA) can be used to find a low-dimensional projection in which the  $G$  groups are well separated<sup>9,19,56</sup>. LDA identifies an ordered set of  $G - 1$  directions in which the between-group variance is maximized relative to the within-group variance. Consider an example with  $D = 2$  neurons and  $G = 2$  groups (Fig. 4b, top). When the data points are projected onto the  $s_1$  axis, the two groups are well separated.

If there are multiple dependent variables for each data point (for example, stimulus identity and decision identity), one might seek to 'demix' the effects of the different dependent variables, such that each projection axis (that is, latent variable) captures the variance of a

single dependent variable. This is often helpful for orienting the user in the low-dimensional space by assigning meaning to the projection axes in terms of externally measurable variables. There are three closely related methods that have been used in the neuroscience literature, which we collectively call demixed dimensionality reduction: a variant of linear regression<sup>4</sup>, a difference of covariances approach<sup>16</sup> and a probabilistic extension<sup>57</sup>. Consider an example with  $D = 2$  neurons and two attributes (dot size and dot color) for each data point (Fig. 4c). Applying demixed dimensionality reduction to these data yields a direction  $s_1$ , which optimally explains the variance in dot color, and an orthogonal direction  $s_2$  (not shown), which explains the variance in dot size. The organization in dot color can be seen when the data points are projected onto  $s_1$ . A similar organization in dot size could be seen by projecting the data onto the orthogonal  $s_2$  axis. Note that the two attributes vary along orthogonal axes (Fig. 4c), although this need not be the case in real data. Methodologically, the variant of linear regression should be used when the dependent variables take on a continuum of values (rather than a few discrete values), whereas the difference of covariances approach should be used when there is no obvious ordering of the values of the dependent variables (for example, different stimulus categories).

### Nonlinear dimensionality reduction methods

Most of the methods presented thus far define a linear relationship between the latent and observed variables (Figs. 2 and 4). In general, the data may lie on a low-dimensional, nonlinear manifold in the high-dimensional space. Depending on the form of the nonlinearity, a linear method may require more latent variables than the number of true dimensions in the data. Two of the most prominent methods to identify nonlinear manifolds are Isomap<sup>79</sup> and locally linear embedding (LLE)<sup>13,14,44,45,58,80</sup>. As with linear methods, the low-dimensional embeddings produced by nonlinear methods should be interpreted with care. Several nonlinear methods use local neighborhoods to estimate the structure of the manifold. Because population responses typically do not evenly explore the high-dimensional space (a problem that grows exponentially with the number of neurons), local neighborhoods might include only temporally adjacent points along the same trajectory. As a result, differences between trajectories can be magnified in the low-dimensional embedding and should be interpreted accordingly. To obtain more even sampling of the high-dimensional space, it will be necessary to substantially increase the richness and diversity of standard task paradigms (for example, the presented stimuli or elicited behavior). Furthermore, nonlinear dimensionality reduction methods are often fragile in the presence of noise<sup>81</sup>, which limits their use in single-trial population analyses. These caveats suggest linear dimensionality reduction as a sensible starting point for most analyses. Before proceeding to nonlinear methods, one should ensure that there is a dense enough sampling of the high-dimensional space such that that local neighborhoods involve data points from different trajectories (or experimental conditions), and that, in the case of single-trial analyses, the nonlinear method is robust to the Poisson-like spiking variability of neurons.

### Practical use

Depending on the scientific question being asked, one should first select an appropriate dimensionality reduction method using the guidelines described above. One can then

perform the necessary data preprocessing (for example, take spike counts, average across trials and/or smooth across time) and apply the selected method to the population activity. This latter step includes finding the latent dimensionality, estimating the model parameters (if applicable), and projecting the high-dimensional data into the low-dimensional space (akin to Fig. 2). This yields a low-dimensional representation of the population activity. This section provides practical guidelines for preprocessing the data, estimating and interpreting the latent dimensionality, running the selected dimensionality reduction method, and visualizing the low-dimensional projections. We point out caveats and potential pitfalls specific to the analysis of population activity, as well as more general pitfalls pertaining to the analysis of high-dimensional data.

### Data preprocessing

The data should be preprocessed to ensure sensible inputs for dimensionality reduction. A few typical pitfalls exist. First, one should ensure that the neurons do not covary as a result of trivial (that is, non-biological) reasons, which can seriously confound any dimensionality reduction method. Examples include electrical cross-coupling between electrodes, which induces positive correlation between neurons<sup>12</sup>, and artificially splitting the response of a single neuron into two (whether as a result of neuron proximity in optical recordings or of spike sorting with electrode recordings), which induces negative correlation between the neurons. Second, neurons with low mean firing rates (for example, less than one spike per second) should typically be excluded, as nearly zero variance for any neuron can lead to numerical instability with some methods. Third, for PCA, one may consider normalizing (that is,  $z$  scoring) the activity of each neuron, as PCA can be dominated by neurons with the highest modulation depths. This is less of a concern for most of the other dimensionality reduction methods, where the latent variables are invariant to the scale of each neuron's activity<sup>61</sup>. Following these considerations, the data can be preprocessed by taking binned spike counts, averaging across trials and/or kernel smoothing across time<sup>33</sup>.

### Estimating and interpreting dimensionality

Many dimensionality reduction methods require a choice of dimensionality ( $K$ ) for the low-dimensional projection. Dimensionality can be thought of as the number of directions explored by the population activity in the high-dimensional firing rate space (Fig. 2). Scientifically, dimensionality is a complexity measure of the population activity and may be suggestive of underlying circuit mechanisms. For example, low dimensionality might suggest that only a small number of common drivers are responsible for the population activity. On the other hand, a higher dimensionality may provide an advantage for downstream neurons reading out information from the recorded population<sup>15,82</sup>.

The most basic approach to estimating dimensionality is to choose a cutoff value for the variance explained by the low-dimensional projection and to choose  $K$  such that the cutoff is exceeded. Given that the cutoff value is often arbitrary, cross-validation may be preferred to ask how many dimensions generalize to explaining held-out data. For probabilistic methods (such as FA, GPFA, LDS and NLDS), one can identify the dimensionality that maximizes the cross-validated data likelihood. Alternatively, for all linear methods and some nonlinear methods, a cross-validated leave-neuron-out prediction error can be computed in the place of

data likelihood<sup>12</sup>. Another approach is to evaluate the number of binary classifications that can be implemented by a linear classifier<sup>15</sup>. In general, the estimated dimensionality can be influenced by the choice of estimation method, by the number of neurons included, by the richness of the experimental setting and by the number of data points in a given data set. These considerations suggest it is safer to make relative, rather than absolute, statements about dimensionality.

### Computational runtime

Although a detailed analysis of the computational runtime for each method is beyond the scope of this review, we will discuss a few rules of thumb. For the linear methods, there are two steps: estimating the model parameters and then projecting the data into the low-dimensional space. For estimating the model parameters, methods (such as PCA and LDA) that require only a single matrix decomposition tend to be faster than methods (such as FA, GPFA, LDS and NLDS) that use an iterative algorithm (such as expectation-maximization) or subspace identification methods<sup>75,76,83</sup>. Methods that involve a dynamics model (such as GPFA, LDS and NLDS) tend to require more computation than those that do not (such as FA). Relative to estimating the model parameters, the second step of projecting the data into the low-dimensional space tends to be fast for all linear methods. To estimate the latent dimensionality, cross-validation tends to be highly computationally demanding because it requires that the model parameters be fit  $m \times n$  times, where  $m$  is the number of cross-validation folds and  $n$  is the number of candidate latent dimensionalities. As the number of recorded neurons continues to grow, computational efficiency will become an increasingly important consideration in the use of dimensionality reduction for neuroscience.

### Visualization

Ideally, one would like to visualize the extracted latent variables directly in the  $K$ -dimensional space. If  $K = 3$ , standard plotting can be used. For larger  $K$ , one possibility is to visualize a small number of two-dimensional projections, which may miss salient features or provide a misleading impression of the latent variables, though tools for rapid visualization exist to help address this limitation<sup>60</sup> (software available at <http://bit.ly/117MTdB>).

### Potential pitfalls when analyzing high-dimensional data

When analyzing multivariate data, it is important to bear in mind that intuition from two- or three-dimensional space may not hold in higher dimensional spaces. For example, one may ask whether two different low-dimensional spaces have a similar orientation in the high-dimensional space, as assessed by comparing angles between vectors in the high-dimensional space. With increasing dimensionality, two randomly chosen vectors will become increasingly orthogonal. Thus, the assessment of orthogonality should be performed relative to the chance distribution of angles rather than an intuitive expectation from lower dimensional spaces. As a second example, the ability of LDA to separate two classes of training data improves as the data dimensionality increases (for a fixed number of data points). With high enough dimensionality, any fixed number of data points can be separated arbitrarily well. These two examples emphasize that caution is needed when analyzing high-dimensional data; many other examples have appeared in the literature<sup>84</sup>.

## Broader connections

We have focused on data contexts in which dimensionality reduction is an appropriate analytical approach. However, dimensionality reduction is by no means the only method available for analyzing neural population activity. For decades, studies have considered how the activity of pairs of neurons covaries<sup>21,85,86</sup>. Moving to larger populations, studies have characterized population activity using regression-based generalized linear models (GLMs), spike word distributions and decoding approaches. Here we describe these related methods, their connection to dimensionality reduction and the contexts in which they can be a more appropriate choice.

First, as the name implies, a GLM is a generalization of the conventional linear-Gaussian relationship between explanatory variables and population activity<sup>87</sup>. A GLM can model a spike train directly using a Poisson count distribution or a point-process distribution. It can be used in settings in which the explanatory variables are observed<sup>64,66</sup> or unobserved (that is, latent)<sup>71,72,74–77</sup>. The former is a generalization of linear regression, whereas the latter is a generalization of linear dimensionality reduction. Thus, the choice of whether to use a GLM (rather than a more conventional linear-Gaussian model) is separate from the choice of whether to perform dimensionality reduction. It is also possible to use a GLM in settings in which there are both observed and unobserved explanatory variables<sup>72,88,89</sup>. GLM are widely used in the regression setting to explain the firing of a neuron in terms of the recent history of the entire recorded population and the stimulus. This approach is appropriate in settings in which one believes that most (or all) of the relevant explanatory variables are observed and can provide insight into stimulus dependence and functional connectivity. In contrast, dimensionality reduction (with or without GLM) should be used if the relevant explanatory variables are unobserved (for example, as a result of common input or unobserved neurons in the network)<sup>74</sup> to address questions about collective population dynamics and variability.

Second, a non-parametric approach for characterizing population activity is to measure the probability of observing every possible spike count vector (termed spike word)<sup>90,91</sup>. This yields a discrete probability distribution, which can then be compared across experimental conditions using information theoretic measures<sup>92</sup>. This method captures higher order correlations across neurons and preserves the precise timing of spikes relative to many dimensionality reduction methods that are based on second-order statistics and the assumption of an underlying firing rate<sup>32</sup>. Whether it is necessary to take into account higher order correlations<sup>90,91</sup> and precise spike times<sup>93</sup> depends on the particular brain region or scientific question being studied. Because the number of possible spike words grows exponentially with the number of neurons, spike word analyses are often limited to a few tens of neurons and require large amounts of data.

Third, dimensionality reduction methods are closely related to population decoding methods. Similar to dimensionality reduction, decoding reduces the high-dimensional population activity to a smaller number of variables. Some prominent examples include decoding sensory stimuli<sup>66,94</sup>, physical location<sup>62</sup>, arm movements<sup>65,95,96</sup>, working memory<sup>97</sup>, object information<sup>98</sup> and more. The key distinction is that decoding seeks to

predict external variables, whereas dimensionality reduction produces low-dimensional representations (latent variables) of the neural activity itself. The fact that decoding performance tends to increase with the number of neurons indicates that a population of neurons can provide more information about either an external or internal variable, compared with a single neuron<sup>99</sup>.

## DISCUSSION

One of the major pursuits of science is to explain complex phenomena in simple terms. Systems neuroscience is no exception, and decades of research have attempted to find simplicity at the level of individual neurons. Standard analysis procedures include constructing simple parametric tuning curves and response fields, analyzing only a select subset of the recorded neurons, and creating population averages by averaging across neurons and trials. Recently, studies have begun to embrace single-neuron heterogeneity and seek simplicity at the level of the population<sup>4,11,15,16</sup>, as enabled by dimensionality reduction. This approach has already provided new insight about how network dynamics give rise to sensory, cognitive and motor function. With the ever-growing interest in studying cognitive and other internal brain processes, along with the continued development and adoption of large-scale recording technologies, dimensionality reduction and related methods may become increasingly essential.

## ACKNOWLEDGMENTS

We are grateful to the authors and publishers of the works highlighted in Figure 3 for permission to reuse portions of their figures. We thank C. Chandrasekaran, A. Miri, W. Newsome, B. Raman, and the members of the laboratories of A. Batista, S. Chase and M. Churchland for helpful discussion during the preparation of this manuscript. This work was supported by the Grossman Center for the Statistics of Mind (J.P.C.), the Simons Foundation (SCGB-325171 and SCGB-325233 to J.P.C.), the Gatsby Charitable Foundation (J.P.C.) and the US National Institutes of Health National Institute of Child Health and Human Development (R01-HD-071686 to B.M.Y.).

## REFERENCES

1. Kipke DR, et al. Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical opportunities. *J. Neurosci.* 2008; 28:11830–11838. [PubMed: 19005048]
2. Kerr JN, Denk W. Imaging *in vivo*: watching the brain in action. *Nat. Rev. Neurosci.* 2008; 9:195–205. [PubMed: 18270513]
3. Ahrens MB, et al. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods.* 2013; 10:413–420. [PubMed: 23524393]
4. Mante V, et al. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature.* 2013; 503:78–84. [PubMed: 24201281]
5. Kaufman MT, et al. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* 2014; 17:440–448. [PubMed: 24487233]
6. Mazor O, Laurent G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron.* 2005; 48:661–673. [PubMed: 16301181]
7. Harvey DC, et al. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature.* 2012; 484:62–68. [PubMed: 22419153]
8. Stokes MG, et al. Dynamic coding for cognitive control in prefrontal cortex. *Neuron.* 2013; 78:364–375. [PubMed: 23562541]
9. Briggman KL, Abarbanel HDI, Kristan WB Jr. Optical imaging of neuronal populations during decision-making. *Science.* 2005; 307:896–901. [PubMed: 15705844]

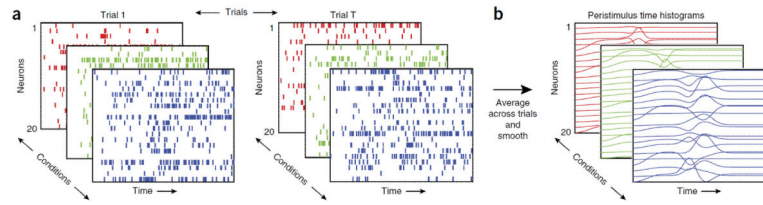
10. Churchland MM, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 2010; 13:369–378. [PubMed: 20173745]
11. Churchland MM, et al. Neural population dynamics during reaching. *Nature.* 2012; 487:51–56. [PubMed: 22722855]
12. Yu BM, et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* 2009; 102:614–635. [PubMed: 19357332]
13. Broome BM, et al. Encoding and decoding of overlapping odor sequences. *Neuron.* 2006; 51:467–482. [PubMed: 16908412]
14. Saha D, et al. A spatiotemporal coding mechanism for background-invariant odor recognition. *Nat. Neurosci.* 2013; 16:1830–1839. [PubMed: 24185426]
15. Rigotti M, et al. The importance of mixed selectivity in complex cognitive tasks. *Nature.* 2013; 497:585–590. [PubMed: 23685452]
16. Machens CK, et al. Functional, but not anatomical, separation of 'what' and 'when' in prefrontal cortex. *J. Neurosci.* 2010; 30:350–360. [PubMed: 20053916]
17. Cohen MR, Maunsell JHR. A neuronal population measure of attention predicts behavioral performance on individual trials. *J. Neurosci.* 2010; 30:15241–15253. [PubMed: 21068329]
18. Luczak A, et al. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron.* 2009; 62:413–425. [PubMed: 19447096]
19. Durstewitz D, et al. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron.* 2010; 66:438–448. [PubMed: 20471356]
20. Bouchard KE, et al. Functional organization of human sensorimotor cortex for speech articulation. *Nature.* 2013; 495:327–332. [PubMed: 23426266]
21. Brown EN, et al. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 2004; 7:456–461. [PubMed: 15114358]
22. Seidemann E, Meilijson I, Abeles M, Bergman H, Vaadia E. Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.* 1996; 16:752–768. [PubMed: 8551358]
23. Jones LM, et al. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proc. Natl. Acad. Sci. USA.* 2007; 104:18772–18777. [PubMed: 18000059]
24. Ponce-Alvarez A, et al. Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. *J. Neurosci.* 2012; 32:11956–11969. [PubMed: 22933781]
25. Horwitz GD, Newsome WT. Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. *J. Neurophysiol.* 2001; 86:2543–2558. [PubMed: 11698541]
26. Bollimunta A, Totten D, Ditterich J. Neural dynamics of choice: single-trial analysis of decision-related activity in parietal cortex. *J. Neurosci.* 2012; 32:12684–12701. [PubMed: 22972993]
27. Sanger TD, Kalaska JF. Crouching tiger, hidden dimensions. *Nat. Neurosci.* 2014; 17:338–340. [PubMed: 24569828]
28. Laurent G. Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.* 2002; 3:884–895. [PubMed: 12415296]
29. Churchland MM, et al. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron.* 2010; 68:387–400. [PubMed: 21040842]
30. Vogels TP, et al. Neural network dynamics. *Annu. Rev. Neurosci.* 2005; 28:357–376. [PubMed: 16022600]
31. Nawrot MP, et al. Measurement of variability dynamics in cortical spike trains. *J. Neurosci. Methods.* 2008; 169:374–390. [PubMed: 18155774]
32. Churchland MM, Abbott LF. Two layers of neural variability. *Nat. Neurosci.* 2012; 15:1472–1474. [PubMed: 23103992]
33. Cunningham JP, et al. Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Netw.* 2009; 22:1235–1246. [PubMed: 19349143]
34. Cox, DR.; Isham, V. *Point Processes.* Chapman and Hall; London: 1980.
35. Tolhurst DJ, Movshon JA, Dean AF. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 1983; 23:775–785. [PubMed: 6623937]



36. Afshar A, et al. Single-trial neural correlates of arm movement preparation. *Neuron*. 2011; 71:555–564. [PubMed: 21835350]
37. Levi R, et al. The role of sensory network dynamics in generating a motor program. *J. Neurosci*. 2005; 25:9807–9815. [PubMed: 16237184]
38. Sasaki T, Matsuki N, Ikegaya Y. Metastability of active ca3 networks. *J. Neurosci*. 2007; 27:517–528. [PubMed: 17234584]
39. Ecker AS, et al. State dependence of noise correlations in macaque primary visual cortex. *Neuron*. 2014; 82:235–248. [PubMed: 24698278]
40. Jun JK, et al. Heterogenous population coding of a short-term memory and decision task. *J. Neurosci*. 2010; 30:916–929. [PubMed: 20089900]
41. Shenoy KV, et al. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci*. 2013; 36:337–359. [PubMed: 23725001]
42. Ames KC, et al. Neural dynamics of reaching following incorrect or absent motor preparation. *Neuron*. 2014; 81:438–451. [PubMed: 24462104]
43. Sadtler PT, et al. Neural constraints on learning. *Nature*. in the press.
44. Stopfer M, Jayaraman V, Laurent G. Intensity versus identity coding in an olfactory system. *Neuron*. 2003; 39:991–1004. [PubMed: 12971898]
45. Brown SL, et al. Encoding a temporally structured stimulus with a temporally structured neural representation. *Nat. Neurosci*. 2005; 8:1568–1576. [PubMed: 16222230]
46. Bathellier B, et al. Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. *Neuron*. 2008; 57:586–598. [PubMed: 18304487]
47. Geffen MN, et al. Neural encoding of rapidly fluctuating odors. *Neuron*. 2009; 61:570–586. [PubMed: 19249277]
48. Raman B, Joseph J, Tang J, Stopfer M. Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J. Neurosci*. 2010; 30:1994–2006. [PubMed: 20147528]
49. Sussillo D, Abbott LF. Generating coherent patterns of activity from chaotic neural networks. *Neuron*. 2009; 63:544–557. [PubMed: 19709635]
50. Ahrens MB, et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*. 2012; 485:471–477. [PubMed: 22622571]
51. Freeman J, et al. Mapping brain activity at scale with cluster computing. *Nat. Methods*. Jul 27.2014 doi:10.1038/nmeth.3041.
52. Nicolelis MA, Baccala LA, Lin RC, Chapin JK. Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science*. 1995; 268:1353–1358. [PubMed: 7761855]
53. Paz R, Natan C, Boraud T, Bergman H, Vaadia E. Emerging patterns of neuronal responses in supplementary and primary motor areas during sensorimotor adaptation. *J. Neurosci*. 2005; 25:10941–10951. [PubMed: 16306407]
54. Lin L, et al. Identification of network-level coding units for real-time representation of episodic experiences in the hippocampus. *Proc. Natl. Acad. Sci. USA*. 2005; 102:6125–6130. [PubMed: 15833817]
55. Narayanan NS, et al. Delay activity in rodent frontal cortex during a simple reaction time task. *J. Neurophysiol*. 2009; 101:2859–2871. [PubMed: 19339463]
56. Bartho P, et al. Population coding of tone stimuli in auditory cortex: dynamic rate vector analysis. *Eur. J. Neurosci*. 2009; 30:1767–1778. [PubMed: 19840110]
57. Brendel W, et al. Demixed principal component analysis. *Adv. Neural Inf. Process. Syst*. 2011; 24:2654–2662.
58. Carrillo-Reid L, et al. Encoding network states by striatal cell assemblies. *J. Neurophysiol*. 2008; 99:1435–1450. [PubMed: 18184883]
59. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. *Cell*. 2006; 125:143–160. [PubMed: 16615896]
60. Cowley BR, et al. Datahigh: graphical user interface for visualizing and interacting with high-dimensional neural activity. *J. Neural Eng*. 2013; 10:066012. [PubMed: 24216250]

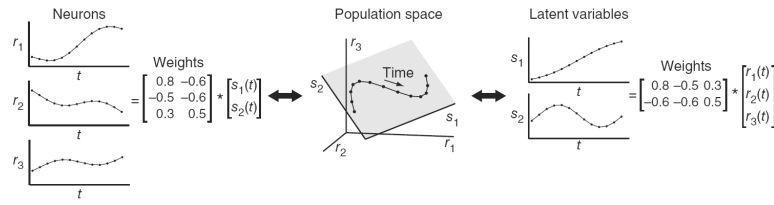
61. Roweis S, Ghahramani Z. A unifying review of linear gaussian models. *Neural Comput.* 1999; 11:305–345. [PubMed: 9950734]
62. Brown EN, et al. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* 1998; 18:7411–7425. [PubMed: 9736661]
63. Eden UT, et al. Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput.* 2004; 16:971–998. [PubMed: 15070506]
64. Truccolo W, et al. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol.* 2005; 93:1074–1089. [PubMed: 15356183]
65. Wu W, et al. Bayesian population decoding of motor cortical activity using a kalman filter. *Neural Comput.* 2006; 18:80–118. [PubMed: 16354382]
66. Pillow JW, et al. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature.* 2008; 454:995–999. [PubMed: 18650810]
67. Shimazaki H, et al. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLOS Comput. Biol.* 2012; 8:e1002385. [PubMed: 22412358]
68. Abeles M, et al. Cortical activity flips among quasi-stationary states. *Proc. Natl. Acad. Sci. USA.* 1995; 92:8616–8620. [PubMed: 7567985]
69. Danoczy M, Hahnloser R. Efficient estimation of hidden state dynamics from spike trains. *Adv. Neural Inf. Process. Syst.* 2006; 8:227–234.
70. Kemere C, et al. Detecting neural-state transitions using hidden markov models for motor cortical prostheses. *J. Neurophysiol.* 2008; 100:2441–2452. [PubMed: 18614757]
71. Smith AC, Brown EN. Estimating a state-space model from point process observations. *Neural Comput.* 2003; 15:965–991. [PubMed: 12803953]
72. Kulkarni JE, Paninski L. Common-input models for multiple neural spike-train data. *Network.* 2007; 18:375–407. [PubMed: 17943613]
73. Paninski L, et al. A new look at state-space models for neural data. *J. Comput. Neurosci.* 2010; 29:107–126. [PubMed: 19649698]
74. Macke JH, et al. Empirical models of spiking in neural populations. *Adv. Neural Inf. Process. Syst.* 2011; 24:1350–1358.
75. Buesing L, Macke J, Sahani M. Spectral learning of linear dynamics from generalized-linear observations with application to neural population data. *Adv. Neural Inf. Process. Syst.* 2012; 25:1691–1699.
76. Pfau D, et al. Robust learning of low-dimensional dynamics from large neural ensembles. *Adv. Neural Inf. Process. Syst.* 2013; 26:2391–2399.
77. Yu BM, et al. Extracting dynamical structure embedded in neural activity. *Adv. Neural Inf. Process. Syst.* 2006; 18:1545–1552.
78. Petreska B, et al. Dynamical segmentation of single trials from population neural data. *Adv. Neural Inf. Process. Syst.* 2011; 24:756–764.
79. Tenenbaum JB, et al. A global geometric framework for nonlinear dimensionality reduction. *Science.* 2000; 290:2319–2323. [PubMed: 11125149]
80. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science.* 2000; 290:2323–2326. [PubMed: 11125150]
81. Boots, B.; Gordon, G. Two-manifold problems with applications to nonlinear system identification. In: Langford, J.; Pineau, J., editors. *Proceedings of the 29th International Conference on Machine Learning.* Omnipress; New York: 2012. p. 623-630.
82. Salinas E, Abbott LF. Vector reconstruction from firing rates. *J. Comput. Neurosci.* 1994; 1:89–107. [PubMed: 8792227]
83. Overschee, PV.; Moor, BD. *Subspace Identification For Linear Systems: Theory, Implementation, Applications.* Kluwer Academic Publishers; 1996.
84. Diaconis P, Freedman D. Asymptotics of graphical projection pursuit. *Ann. Stat.* 1984; 12:793–815.

85. Gerstein GL, Perkel DH. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science*. 1969; 164:828–830. [PubMed: 5767782]
86. Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 2011; 14:811–819. [PubMed: 21709677]
87. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. Vol. 37. Chapman and Hall; 1998.
88. Lawhern V, et al. Population decoding of motor cortical activity using a generalized linear model with hidden states. *J. Neurosci. Methods*. 2010; 189:267–280. [PubMed: 20359500]
89. Vidne M, et al. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *J. Comput. Neurosci.* 2012; 33:97–121. [PubMed: 22203465]
90. Shlens J, et al. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 2006; 26:8254–8266. [PubMed: 16899720]
91. Schneidman E, et al. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*. 2006; 440:1007–1012. [PubMed: 16625187]
92. Berkes P, et al. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*. 2011; 331:83–87. [PubMed: 21212356]
93. Kumar A, Rotter S, Aertsen A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* 2010; 11:615–627. [PubMed: 20725095]
94. Stanley GB, Li FF, Dan Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* 1999; 19:8036–8042. [PubMed: 10479703]
95. Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. *Science*. 1986; 233:1416–1419. [PubMed: 3749885]
96. Gilja V, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* 2012; 15:1752–1757. [PubMed: 23160043]
97. Baeg EH, et al. Dynamics of population code for working memory in the prefrontal cortex. *Neuron*. 2003; 40:177–188. [PubMed: 14527442]
98. Hung CP, et al. Fast readout of object identity from macaque inferior temporal cortex. *Science*. 2005; 310:863–866. [PubMed: 16272124]
99. Quiroga RQ, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* 2009; 10:173–185. [PubMed: 19229240]
100. Santhanam G, et al. Factor-analysis methods for higher-performance neural prostheses. *J. Neurophysiol.* 2009; 102:1315–1330. [PubMed: 19297518]

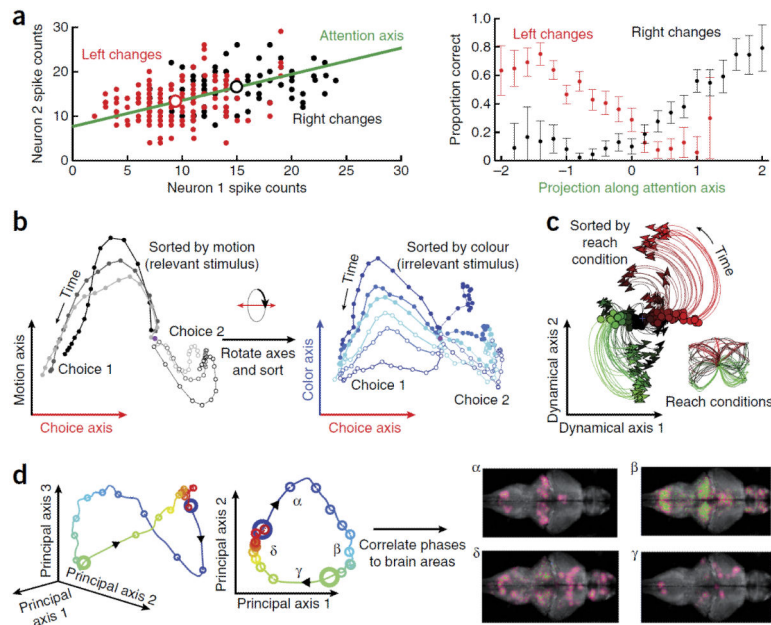


**Figure 1.**

Motivation for population analyses and dimensionality reduction. **(a)** Enabled by the growth in scale and resolution of neural recording technologies, a typical experiment yields a collection of many trials (sets of panels from left to right), many experimental conditions (different colored panels shown in depth) and many neurons (rows of each panel, shown as spike rasters). Scrutinizing these data qualitatively and quantitatively presents many challenges, both for basic understanding and for testing hypotheses. **(b)** Neural responses are often averaged across trials (within a given condition) and smoothed into a peristimulus time histogram. Even these trial-averaged views can be difficult to interpret, as the number of conditions and recorded neurons grows. Notably, this challenge can be present even in data with simple structure: each of these simulated neurons have Poisson spiking with an underlying firing rate that is a windowed linear mixture of three Gaussian pulses. Each neuron has different mixture coefficients, baseline and amplitude. Dimensionality reduction is one class of statistical methods that can extract simple structure from these seemingly complex data.

**Figure 2.**

Conceptual illustration of linear dimensionality reduction for three neurons ( $D = 3$ ) and two latent variables ( $K = 2$ ). Center, the population activity (black points) lies in a plane (shaded gray). Each point represents the population activity at a particular time and can be equivalently referred to using its high-dimensional coordinates  $[r_1, r_2, r_3]$  or low-dimensional coordinates  $[s_1, s_2]$ . The points trace out a trajectory over time (black curve). Left, the population activity  $r_1, r_2$  and  $r_3$  can be reconstructed by taking a weighted combination of the latent variables, where the weights are specified by the matrix shown. Right, the latent variables  $s_1$  and  $s_2$  can be obtained by taking a weighted combination of the population activity, where the weights are specified by the matrix shown.

**Figure 3.**

Examples of scientific studies using dimensionality reduction. **(a)** Single-trial statistical power, visual attention. Left, a didactic example of projecting the responses of two neurons onto the attention axis (green; units, spike counts). For a V4 population (right), the normalized position (that is, projection) along this attention axis was predictive of behavior on single trials: the farther the projection was to the upper right of the attention axis, the more likely the animal was of correctly detecting right changes and the less likely the animal was of correctly detecting left changes. Adapted with permission from ref. 17. **(b)** Population response structure, decision-making. The population activity recorded in prefrontal cortex was projected onto three axes (units, spikes per s): the axis of evidence integration (choice axis), the relevant stimulus axis (motion axis) and the irrelevant stimulus axis (color axis). Each trace corresponds to responses averaged across trials of the same dot motion (gray traces) or dot color (blue traces). Despite the apparent complexity of single-neuron responses, the population activity shows orderly structure across different conditions of dot motion and color and suggests a network mechanism for gating and integration of information in prefrontal cortex. Adapted with permission from ref. 4. **(c)** Population response structure, motor system. The population activity recorded in motor cortex was projected onto a plane (units, spikes per s) where simple (rotational) dynamics are best captured; different traces are different experimental conditions (arm reaches shown with the same color in the inset). Dots denote the preparatory (pre-movement) neural activity, suggesting a mechanistic explanation for single neuron complexity: preparatory responses set the initial state of a population-level dynamical system that runs through movement. Adapted with permission from ref. 11. **(d)** Exploratory data analysis, brain-wide. The population activity recorded throughout the brain of larval zebrafish was projected onto its principal components for visualization purposes (the same data is shown at left plotted in three and two dimensions, units of  $F/F$ ). Four phases of response were identified (labeled  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ ), which were then connected back to distinct neural structures. Right, axial

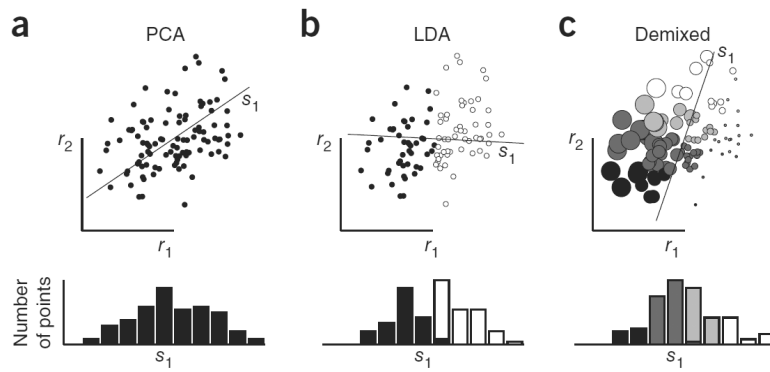
views during each phase, where green dots indicate active neurons with magenta confidence intervals (caudal-rostral is left-right). Adapted with permission from ref. 50.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4.** Conceptual illustration of PCA, LDA and demixed dimensionality reduction for two neurons ( $D = 2$ ). **(a)** PCA finds the direction ( $s_1$  axis) that captures the greatest variance in the data (black dots, top), shown by the projection onto the  $s_1$  axis (bottom). **(b)** LDA finds the direction ( $s_1$  axis) that best separates the two groups of points (black and white dots, top). The separation can be seen in the projection onto the  $s_1$  axis (bottom). **(c)** Demixed dimensionality reduction (using the method described in ref. 16 finds the direction that explains the variance in dot color ( $s_1$  axis, top) and an orthogonal direction ( $s_2$  axis, not shown) that explains the variance in dot size. The organization in dot color can be seen in the projection onto the  $s_1$  axis (bottom). Note that these illustrations were created using the same data points (dots), and it is the use of different methods (which exploit different data features, such as group membership in **(b)** or color and size in **(c)**) that produce different directions  $s_1$  across the top panels and different projections across the bottom panels.



**Table 1**

Overview of dimensionality reduction methods commonly applied to neural population responses

Method	Analysis objective	Temporal smoothing	Explicit noise model	Representative uses in neuroscience (refs.)
PCA	Covariance	No	No	6,29,50
FA	Covariance	No	Yes	10,43,100
LDS/GPFA	Dynamics	Yes	Yes	12,71,72
NLDS	Dynamics	Yes	Yes	77,78
LDA	Classification	No	No	9,19,56
Demixed	Regression	No	Yes/No	4,16
Isomap/LLE	Manifold discovery	No	No	14,44,58

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript