

Comparing high-dimensional neural recordings by aligning their low-dimensional latent representations

Max Dabagia*, Konrad P. Kording[†], and Eva L. Dyer[‡]

Abstract

Many studies in neuroscience rely on modeling changes in neural activity in response to a specific stimulus or perturbation. Quantifying changes in neural activity is challenging, especially when trying to model interactions that occur at the population level or across many neurons. As many analyses involve comparisons across different sets of neurons, a critical requirement for such an approach is that it is somehow invariant to the specific set of neurons drawn from the population and doesn't require one-to-one correspondence. If the activity of a large collection of neurons can be modeled as being driven by a smaller number of *latent factors*, then distinct neural recordings acquired from neurons that share the similar underlying latent space can be compared by finding and aligning their latent factors. In this review, we provide an overview of how latent space models can be generated, aligned, and compared. Being able to align latent representations promises meaningful ways of comparing high-dimensional neural activities across times, subsets of neurons, or subjects.

Introduction

The study of brain function and cognition relies on the measurement and interpretation of changes to the brain's activity when they exist. Whether between states of disease and health [1], across novice- and expert-level performance on a complex task [2], or in wakefulness versus sleep [3], we need ways to compare neural activity patterns and draw conclusions about how they differ.

Historically, changes in neural activity patterns have been examined at the level of single cells, by characterizing how the firing rate (or tuning curves) of each cell shifts across conditions

*MD is in the School of Electrical and Computer Engineering (ECE) at the Georgia Institute of Technology.

[†]KPK is in the Depts of Bioengineering and Neuroscience at the University of Pennsylvania.

[‡]ELD is in the Coulter Department of Biomedical Engineering at the Georgia Institute of Technology.

Corresponding author: evadye@gatech.edu

[4, 5, 6]. However, we know that the brain is a highly interconnected network of neurons [7], and thus the impact of different disease states or conditions is likely to have complex effects across an entire circuit. Without modeling these higher-order interactions between neurons, it may be impossible to model the collective impact of a given perturbation.

Although improving technology allows simultaneous access to an ever-increasing number of neurons [8, 9, 10], each recording is still restricted to a fraction of all the neurons in a circuit of interest. In most preparations, maintaining contact with specific neurons across two recording sessions is challenging or impossible, especially when studying brain activity from two individuals. At first glance, it is tempting to try to pair off similar neurons across recordings and find “correspondence” between specific neurons in each dataset. However, it is unclear whether this is a fruitful approach in most cases (although it has a remarkably successful history in simpler nervous systems, such as those of insects), or whether we should expect to find correspondences at the level of individual units.

The alternative is to summarize each recording in a way that is invariant to the specific subset of neurons that are polled. Many existing approaches build invariance through the computation of metrics that capture essential properties of the pairwise interactions between neurons (like clustering or density of connections) [11, 12, 13], or extract a global property of the network such as criticality [14, 15]. While these approaches can be informative when we know the right metrics to compare across conditions, having a data-driven approach to identify correlations across multiple units would provide a rich multi-dimensional picture of changes in neural circuits.

One potential approach is to first find a simpler representation of the data, or a factorization, and then find correspondence between the factors of this model (rather than individual neurons). Fortunately, neural activity can be approximated by such *latent representations* in many models and systems [16, 17]. A substantial body of evidence shows that these latent factors are stable across subsets of neurons, across time, and even individuals, and may not be obvious in the activity of lone units, suggesting they may be a powerful tool for comparison [18, 19, 20, 21].

In this review, we discuss strategies for comparing high-dimensional neural activity by leveraging low-dimensional latent spaces present across multiple recordings. We then present the challenges that will need to be overcome to compare the recordings at larger scales and across more severe perturbations, and conclude with an overview of the prospective analyses and discoveries these comparison strategies will make possible.

Latent state space models for neural populations

To build a population-level view, the activity of all measured neurons can be represented as a point in a state space, where each axis represents the firing rate of one of the recorded neurons (Figure 1A) [22]. As a result, the dimension of this space is the same as the number of neurons sampled in the recording. If every neuron operated completely independently this would be the end of the story, and moreover neuroscience would be in trouble, because understanding a neural circuit would require the characterization of every single neuron involved in it. Fortunately, there is some evidence that this is not the case. For one thing, neurons interact with each other through both excitation and inhibition, which naturally leads to correlation in their activity [23, 24]. When many neurons are correlated with each other, it is reasonable to expect that their activities might be explainable by a smaller number of variables.

To discover these low-dimensional factors that may drive neural activity, *dimensionality reduction* methods can be used to convert observations of a large number of observed neurons into a smaller number of latent variables. This identifies redundancies (such as correlation) across many measured neural traces in a dataset. The simplest methods identify each latent variable with a particular pattern of neural activity (Figure 1B), such that that any observation can be approximated as a linear combination of these factors [25, 2, 26]. Principal Components Analysis (PCA) is probably the most widely-used technique [20, 27, 28], and chooses latent variables which retain the most variance of the observations possible. There are many extensions of this approach that either relax PCA's assumptions (e.g. the latent variables must be orthogonal) [29], or incorporate additional structure into the solution (e.g., adding a non-negativity constraint) [30, 31]. To leverage temporal information, a variety of methods model the latent factors as a dynamical system and find mappings which preserve coherent dynamics [32, 33, 18, 34]. More sophisticated manifold learning strategies can identify low-dimensional structures embedded in non-linear ways in neural activity, and “flatten” them to recover the latent factors [35, 36].

The need for alignment and stitching

An important feature of latent representations is that they summarize neural recordings in a way that (in theory) is sensitive to only the factors driving the activity of a population, and invariant to which particular neurons are tracked. However, in practice, both the dimensionality reduction technique used to generate the latent representation and the reality of having access to only a fraction of all the neurons in a population can transform latent representations, making it impossible to compare them directly without corrective action. For example, if we have two noisy versions of the same dataset (say two recordings from the same neurons), the

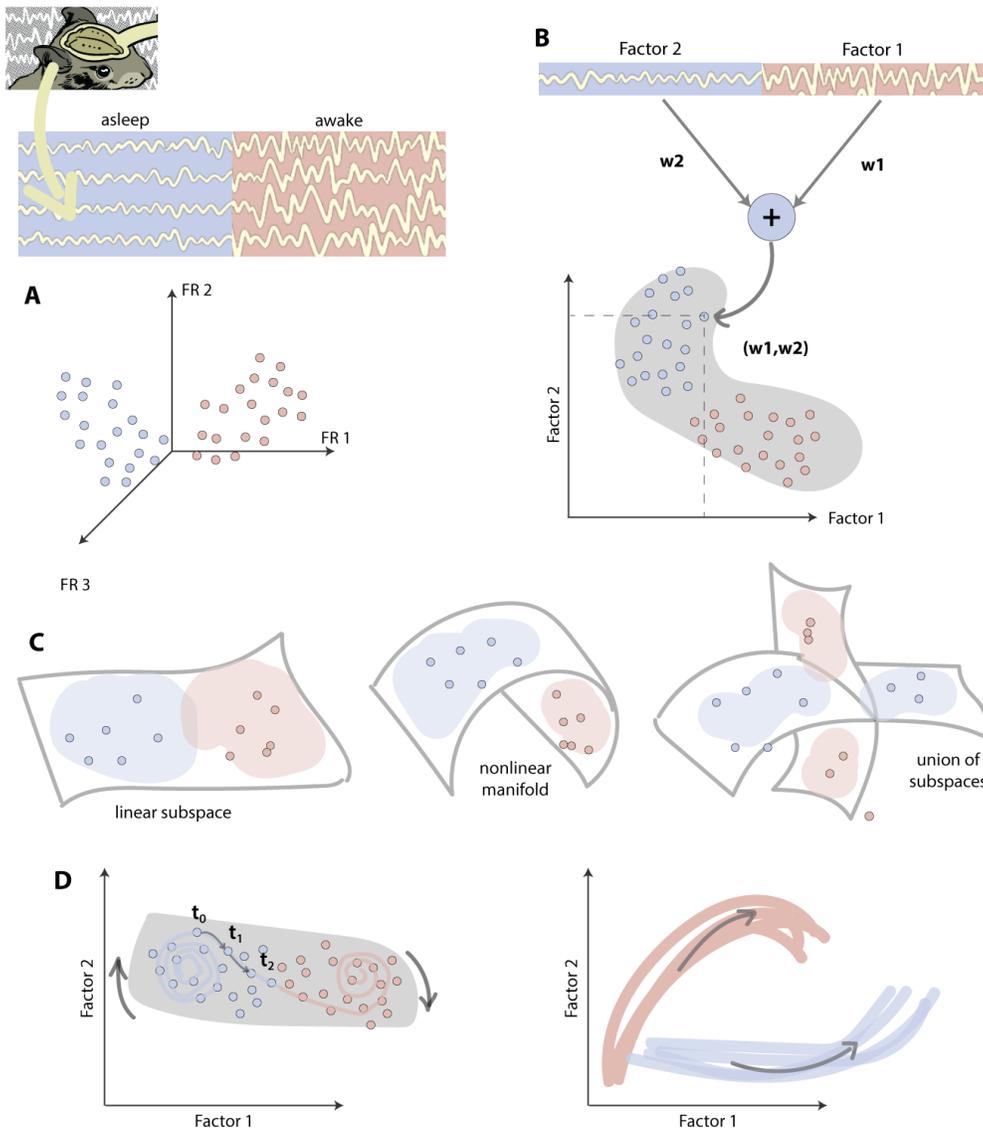


Figure 1: *Low-dimensional and dynamical latent space models.* In A, we depict two multi-unit neural recordings collected in different behavioral states (Left and Right). Below, a visualization of the firing rates of three units, where each axis corresponds to the firing rate of a different neuron. When we record from the same unit across different time points, the changes in firing rate can be observed in the collective firing rates. In B, we depict the formation of a latent space model which takes high-dimensional neural recordings and maps them into a 2D latent space. In C, from left to right, we show a linear subspace, nonlinear manifold, and union of subspaces. Along the bottom row in D, we depict a dynamical system model with the low-dimensional dynamics constraining the possible evolution of states.

order of principal components often changes. Next, we explore strategies for comparing and aligning latent spaces across recordings, both across time points and also across potentially different subsets of neurons.

Before proceeding, however, we will describe in further detail the obstacles to directly comparing latent representations. PCA, for instance, yields a set of latent factors ordered by how much variance they explain, but other linear methods find factors which are chosen and ordered more arbitrarily, typically unique only up to an overall rotation, and the situation may only worsen with nonlinear methods. Another possibility is that the particular subset of neurons might be significantly different; for example, if one recording contains a higher proportion of neurons strongly correlated with a particular latent variable, the relative importance of that latent variable will be inflated. Much as in the previous case, this type of problem results in a transformation of the latent representation that is not difficult to reverse once it is identified, but challenging to predict and absolutely necessary to account for in order to allow comparison. Thus, our task is to find a transformation that maps the two embeddings into the same space such that they align with one another.

Distribution alignment. To tackle the issue of finding correspondence, or a way to compare two latent representations, one can pose this problem as finding a mapping between two datasets through distribution alignment (DA) (Figure 2A). In this setting, we define a class of transformations to explore and a divergence measure, and then look for a transformation that minimizes the divergence between the transformed source and target, effectively aligning them. A few different approaches for DA have been introduced, tailored to the neural decoding setting. The first, Distribution Alignment Decoding (DAD) [38], uses density estimation to infer the probability distributions in latent space, and then, owing to the non-convexity of Kullback-Leibler divergence, employs a brute-force search to identify the optimal rotation which matches the two distributions. The second, Hierarchical Wasserstein Alignment (HiWA) [37] improves upon this strategy by using the structure of the data to decompose the alignment problem, namely the tendency of neural circuits to constrain their low-dimensional activity to clusters or low-dimensional subspaces. In conjunction with a more geometric formulation of distance drawn from optimal transport, HiWA is able to more quickly and robustly recover the correct rotation. Most recently, the potential of an online alignment strategy to enhance the stability of a BCI controller has been demonstrated [39], which assumes a sufficiently substantial fraction of neurons are stable so that only a few coordinate axes of the neural manifold need to be adjusted. Although less general than distribution alignment, the decrease in computational requirements makes this solution more practical in a realistic online BCI setting. However, all of these nascent distribution alignment approaches are only capable of dealing with the unitary transformations that arise from the dimensionality reduction process. In the case that the

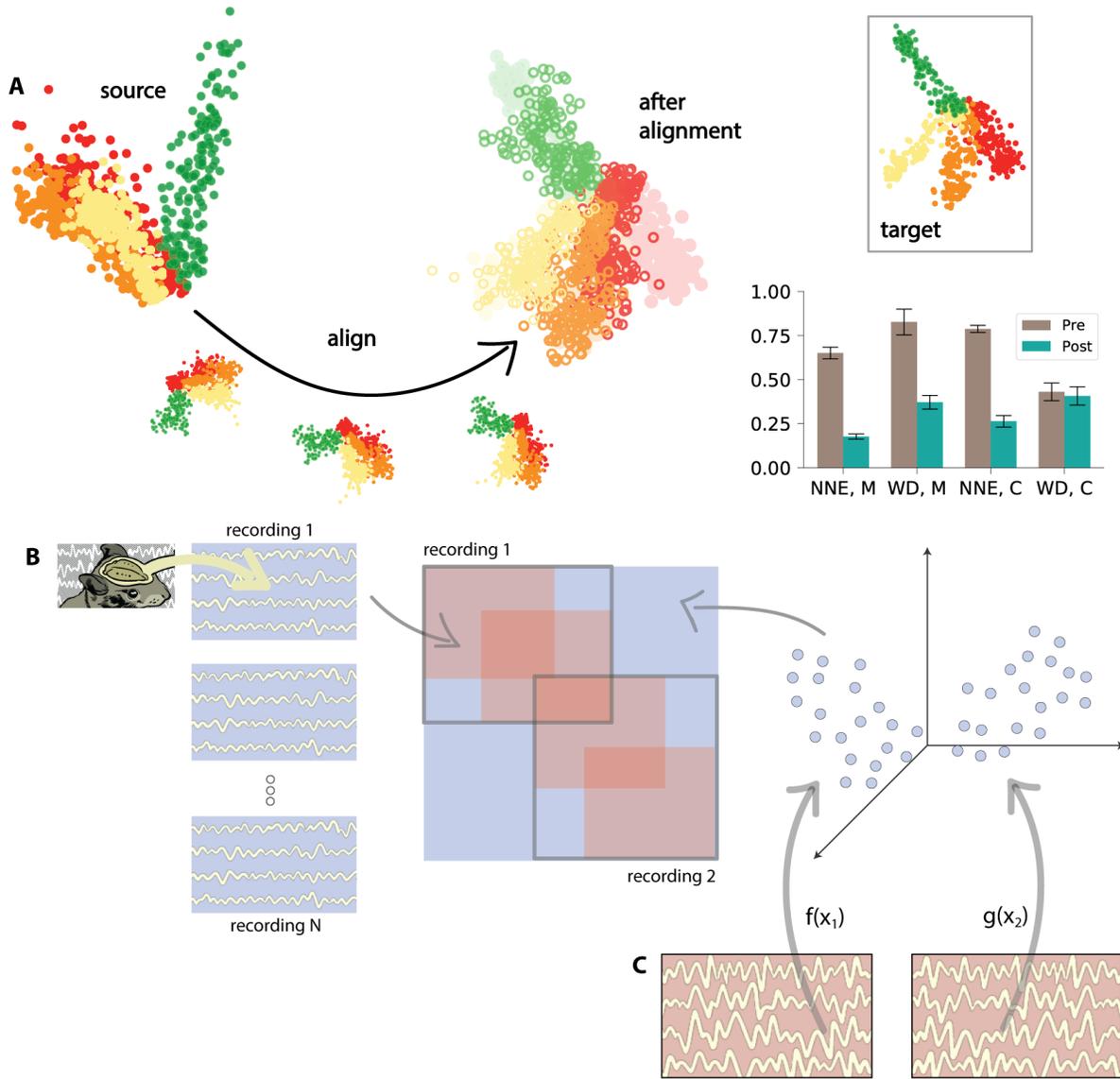


Figure 2: Overview of different alignment methods for neural activity. In *A*, a distribution alignment technique (HiWA [37]) is used to align one latent representation (source) to another structurally-similar one (target). The bar graph (lower right) shows improved cluster prediction accuracy based on the nearest neighbors of each source point within the labelled target. In *B*, the covariance inference approach is demonstrated; the objective is to complete the covariance matrix of the aggregated population of several recordings with neuronal or temporal overlap, which then allows construction of a common, covariance-based latent embedding. In *C*, coupled latent representations $f(x_1)$ and $g(x_2)$ are learned so that the resulting space is shared across the two recordings, which is the strategy of learning coupled mappings.

low-dimensional dynamics of neural circuits are intrinsically different, more sophisticated and nonlinear approaches will be required to match the latent spaces.

A deep learning-driven solution to this problem [40] draws inspiration from generative adversarial networks (GANs), which have received some attention in the computational neuroscience community both for generative tasks like producing realistic neural activity [41, 42] and more novel applications, like decoding visual stimuli from fMRI scans [43]. In essence, the key innovation of GANs is that a network can be indirectly trained to sample from some distribution by instead training it to fool an observer into believing that the samples it generates came from the target distribution. To use this idea for distribution alignment, the generator just learns a transformation of neural activity which aligns it to the distribution of the training set. This approach is promising for brain-machine interfaces, where the priority is just correctly decoding target variables and less about interpreting the embedding. On the other hand, when using latent representations to study a neural circuit, it might be desirable for distinctive new features of the shifted distribution to be preserved, but a sufficiently powerful alignment network which has “overfit” to the target dataset will attenuate them in favor of matching the original dataset as closely as possible.

Covariance inference. An alternative approach for combining recordings is to treat the situation as a missing data problem; that is, the combined pool of all of the neurons from the recordings are assumed to come from the same population, and the challenge is to infer the covariance matrix of the entire population based only on incomplete samples of it [44, 12] (Figure 2B). Normally, finding the population’s latent space would only be possible with overlapping samples, namely those where a number of neurons at least the size of the dimensionality of the latent space are known to be the same. However, by leveraging the dynamics of the population - represented here through time-lagged covariance, or essentially the amount of information gained about the instantaneous firing rate of one neuron by observing the history of another - the covariance matrix can be inferred even with no overlap across samples, as long as the recordings are performed consecutively [11], meaning the temporal gap between the end of one recording and the beginning of the next is too short for the system’s position in state space to shift significantly. These requirements limit this approach’s usefulness, especially in the acute paradigm where no neurons can be reliably matched across recordings and recordings are usually not made serially. Nevertheless, in many datasets (imaging in particular) such recording structure can be leveraged to recover the latent representations, and so this method is a useful counterpart to the more general distribution alignment approach.

Learning independent mappings into a common latent space. The final strategy is to join the process of dimensionality reduction across recordings such that the resulting latent representations are all embedded in the same space (Figure 2C). A straightforward approach

for doing so that has withstood the test of time is canonical correlation analysis (CCA) [45, 46, 47], and more sophisticated nonlinear extensions [48, 49, 50]. This technique is actually a generalization of principal component analysis, where rather than choosing a single component at a time which maximizes variance, it chooses one component for *each dataset* so that the covariance along these components is maximized. CCA reduces to PCA when the datasets being compared are the same. The problem is that if the recordings were not made simultaneously (or aligned to a stimulus), the correspondence of individual moments in time across the datasets is essentially arbitrary, and so the notion of covariance is poorly defined.

A recent method for learning nonlinear dynamical systems (LFADS) “stitches” together data from many recording sessions by learning a separate input and output mapping between each session and the shared latent space for all sessions jointly [51]. This method initializes the mappings between each recording session (dataset) and the latent space by first finding the trial-averaged responses for each unit and then learning a weight matrix (input map) to match these trial-averaged responses. This means that given training data consisting of the direction (or target) the reach is towards, consistency does not need to be enforced across neurons, but rather reaches towards similar targets across days can be aggregated. The result is that task information can be leveraged to stitch together dynamics from different populations of neurons. In this case, a common latent space model is learned across all of the measured units over all days and used to decode on any one of the days.

Challenges

Discovery of latent representations that capture the structures within different neural datasets is critical for the success of alignment. However, using dimensionality reduction to study neural population dynamics can be daunting because it is unsupervised and the resulting latent factors can be hard to interpret. Even with supervised labels or training data (like movement kinematics or sleep state in the examples presented in this review), it can be hard to know when a latent space model is sufficient to describe the structure of the data. It is unclear when latent variables should be considered intrinsic to the functionality of a neural population, or when they are just a concise description of activity. For example, it is not understood whether the ability to decode movement patterns from activity in motor cortex is because neurons have been found which are designed to encode movement, or because their activity is highly correlated with movement as a side-effect of their true functional roles. Most dimensionality reduction techniques are agnostic to this distinction, but through careful design of experiments, theory, and methods, latent representations may be examined and understood in the context of underlying circuit or systems-level mechanisms.

The methods described in the previous section assume that there is some persistent low-dimensional structure that can be observed across two recordings. However, changes in under-

lying brain states and sources of modulation like attention and engagement [52, 53, 54, 55], as well as shift as units die or electrodes move [56, 57, 58]. To overcome these challenges, some supervised data or labels can be provided to the methods to help ground alignment across datasets when the shape of the two distributions is not sufficient to find a common space for the datasets. Additionally, methods for partial domain adaptation and alignment may be utilized to facilitate matching in these conditions. Methods that can account for realistic shifts in neural data will allow for a comparison of data while also modeling state-dependent changes like attention that may modulate the neural distribution across recordings.

As the algorithms to perform these types of analysis grow more sophisticated and specialized, it will be important to have neural datasets to benchmark against that are generated across sufficiently diverse conditions. In machine learning, common datasets used across the community have often spurred progress, from simple benchmarks like MNIST [59] to more modern and powerful tools like the ImageNet dataset [60]. In neuroscience, open-access brain imaging datasets and tools like those provided in the Allen Institute’s optophysiology-based Brain Observatory [61] and their more recent high-density survey with Neuropixel probes [62], can be used to build a common accepted benchmark dataset to compare methods, as well as fertile ground for more in-depth studies. Moving forward, widespread data sharing like this will allow for more reproducible results and pave the way for more powerful and generalizable algorithms.

Outlook

While understanding of the physiological impact of neurological and neurodegenerative diseases has advanced tremendously in recent years, a description of the disruptions they cause to the functionality of neural circuits has not yet been possible. It is highly likely that when such a description is made, it will be through the study of shifts in the activity of many neurons in diseased brains. The alignment strategies outlined in this review and more sophisticated methods built on them will provide the tools necessary to effectively compare the latent spaces that the brain occupies in health and provide signatures of decline as it progresses into disease.

In current brain machine interfaces (BMIs), decoders rely on stable mappings from neural activity to the variables to be decoded, which often degrade over time [63]. Thus with the ability to directly compare latent representations and align them, it will be possible to correct for changes in neural distributions over time to develop more robust neural decoding interfaces. These controllers might be able to adapt to shifts in neural activity due to cognitive state or adaptation more effectively, and offer more sustained long-term performance without collection of new training data. We emphasize that latent space alignment should not be considered only a convenience for analysis, but also a tool for improving the weaknesses of current BMIs and neural prosthetics.

Many of the state-of-the-art approaches to comparison we discussed in this review deal primarily with linear methods for estimating the latent factors. However, there is reason to believe that latent structure may be embedded non-linearly, due to the inherent constraints biology places upon neural activity. In these cases, unions of subspaces [64, 65] and manifolds [66] can be used to capture the low-dimensional structures present in these data. By coupling this model with matrix completion methods [67, 68], it is possible to stitch together manifolds from many brain areas and start to fill in the blanks across recordings that are collected from different sites of contact in the brain (non-overlapping populations). This would provide a useful way to start to compare whole brain recordings even when individual recordings provide only a small view of the bigger picture.

Finally, these alignment strategies could be used to combine data from many subjects into a common latent space. Most of the methods we discussed are still limited to a single individual, often trained on a highly constrained behavior. By expanding the robustness of latent space models and alignment approaches, and their flexibility across contexts, this can allow many more individuals to be aligned in the same common reference space. With the ability to align neural datasets across many individuals into a common latent space, unprecedented comparative analyses would be possible.

References

- [1] Samuel Frere and Inna Slutsky. Alzheimer’s disease: from firing instability to homeostasis network collapse. *Neuron*, 97(1):32–58, 2018.
- [2] Patrick T Sadtler, Kristin M Quick, Matthew D Golub, Steven M Chase, Stephen I Ryu, Elizabeth C Tyler-Kabara, Byron M Yu, and Aaron P Batista. Neural constraints on learning. *Nature*, 512(7515):423, 2014.
- [3] Keith B Hengen, Alejandro Torrado Pacheco, James N McGregor, Stephen D Van Hooser, and Gina G Turrigiano. Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake. *Cell*, 165(1):180–191, 2016.
- [4] Alex Roxin, Nicolas Brunel, David Hansel, Gianluigi Mongillo, and Carl van Vreeswijk. On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience*, 31(45):16217–16226, 2011.
- [5] Logan Chariker, Robert Shapley, and Lai-Sang Young. Orientation selectivity from very sparse LGN inputs in a comprehensive model of macaque V1 cortex. *Journal of Neuroscience*, 36(49):12368–12384, 2016.

- [6] Carsen Stringer, Marius Pachitariu, Nicholas A Steinmetz, Michael Okun, Peter Bartho, Kenneth D Harris, Maneesh Sahani, and Nicholas A Lesica. Inhibitory control of correlated intrinsic variability in cortical networks. *Elife*, 5:e19695, 2016.
- [7] Shreya Saxena and John P Cunningham. Towards the neural population doctrine. *Current Opinion in Neurobiology*, 55:103–111, 2019.
- [8] Ian H Stevenson and Konrad P Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139, 2011.
- [9] Misha B Ahrens, Michael B Orger, Drew N Robson, Jennifer M Li, and Philipp J Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, 10(5):413, 2013.
- [10] Nicholas A Steinmetz, Christof Koch, Kenneth D Harris, and Matteo Carandini. Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, 50:92–100, 2018.
- [11] Daniel Soudry, Suraj Keshri, Patrick Stinson, Min-hwan Oh, Garud Iyengar, and Liam Paninski. Efficient "shotgun" inference of neural connectivity from highly sub-sampled activity data. *PLoS computational biology*, 11(10):e1004464, 2015.
- [12] Marcel Nonnenmacher, Srinivas C Turaga, and Jakob H Macke. Extracting low-dimensional dynamics from multiple large-scale neural population recordings by learning to predict correlations. In *Advances in Neural Information Processing Systems*, pages 5702–5712, 2017.
- [13] Braden AW Brinkman, Fred Rieke, Eric Shea-Brown, and Michael A Buice. Predicting how and when hidden neurons skew measured synaptic interactions. *PLoS computational biology*, 14(10):e1006490, 2018.
- [14] Dante R Chialvo. Emergent complex neural dynamics. *Nature physics*, 6(10):744–750, 2010.
- [15] Zhengyu Ma, Gina G Turrigiano, Ralf Wessel, and Keith B Hengen. Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron*, 104(4):655–664, 2019.
- [16] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500, 2014.
- [17] Matthew R Whiteway and Daniel A Butts. The quest for interpretable models of neural population activity. *Current Opinion in Neurobiology*, 58:86–93, 2019.

- [18] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51, 2012.
- [19] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78, 2013.
- [20] Matthew T Kaufman, Mark M Churchland, Stephen I Ryu, and Krishna V Shenoy. Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440, 2014.
- [21] Ryan C Williamson, Brent Doiron, Matthew A Smith, and Byron M Yu. Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current Opinion in Neurobiology*, 55:40–47, 2019.
- [22] Liam Paninski, Yashar Ahmadian, Daniel Gil Ferreira, Shinsuke Koyama, Kamiar Rahnema Rad, Michael Vidne, Joshua Vogelstein, and Wei Wu. A new look at state-space models for neural data. *Journal of computational neuroscience*, 29(1-2):107–126, 2010.
- [23] Ehud Zohary, Michael N Shadlen, and William T Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140, 1994.
- [24] Larry F Abbott and Peter Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.
- [25] Misha Tsodyks, Tal Kenet, Amiram Grinvald, and A Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286(5446):1943–1946, 1999.
- [26] Artur Luczak, Bruce L McNaughton, and Kenneth D Harris. Packet-based communication in the cortex. *Nature Reviews Neuroscience*, 16(12):745–755, 2015.
- [27] Christopher D Harvey, Philip Coen, and David W Tank. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62, 2012.
- [28] Benjamin R Cowley, Matthew A Smith, Adam Kohn, and M Yu Byron. Stimulus-driven population activity patterns in macaque primary visual cortex. *PLoS computational biology*, 12(12):e1005185, 2016.

- [29] Dennis Child. *The essentials of factor analysis*. Cassell Educational, 1990.
- [30] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.
- [31] Aapo Hyvärinen. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110534, 2013.
- [32] Peiran Gao, Eric Trautmann, Byron M Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.
- [33] Dmitry Kobak, Jose L Pardo-Vazquez, Mafalda Valente, Christian K Machens, and Alfonso Renart. State-dependent geometry of population activity in rat auditory cortex. *eLife*, 8:e44526, 2019.
- [34] Bingni W Brunton, Lise A Johnson, Jeffrey G Ojemann, and J Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1 – 15, 2016.
- [35] Mark Stopfer, Vivek Jayaraman, and Gilles Laurent. Intensity versus identity coding in an olfactory system. *Neuron*, 39(6):991–1004, 2003.
- [36] Elad Ganmor, Ronen Segev, and Elad Schneidman. A thesaurus for a neural population code. *Elife*, 4:e06134, 2015.
- [37] John Lee, Max Dabagia, Eva Dyer, and Christopher Rozell. Hierarchical optimal transport for multimodal distribution alignment. In *Advances in Neural Information Processing Systems*, pages 13453–13463, 2019.
- [38] Eva L Dyer, Mohammad Gheshlaghi Azar, Matthew G Perich, Hugo L Fernandes, Stephanie Naufel, Lee E Miller, and Konrad P Körding. A cryptography-based approach for movement decoding. *Nature Biomedical Engineering*, 1(12):967, 2017.
- [39] Alan D Degenhart, William E Bishop, Emily R Oby, Elizabeth C Tyler-Kabara, Steven M Chase, Aaron P Batista, and M Yu Byron. Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nature Biomedical Engineering*, pages 1–14, 2020.

- [40] Ali Farshchian, Juan A Gallego, Joseph P Cohen, Yoshua Bengio, Lee E Miller, and Sara A Solla. Adversarial domain adaptation for stable brain-machine interfaces. *arXiv preprint arXiv:1810.00045*, 2018.
- [41] Takafumi Arakaki, Gregory Barello, and Yashar Ahmadian. Capturing the diversity of biological tuning curves using generative adversarial networks. *arXiv preprint arXiv:1707.04582*, 2017.
- [42] Manuel Molano-Mazon, Arno Onken, Eugenio Piasini, and Stefano Panzeri. Synthesizing realistic neural population activity patterns using generative adversarial networks. *arXiv preprint arXiv:1803.00338*, 2018.
- [43] Ghislain St-Yves and Thomas Naselaris. Generative adversarial networks conditioned on brain activity reconstruct seen images. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1054–1061. IEEE, 2018.
- [44] Srini Turaga, Lars Buesing, Adam M Packer, Henry Dagleish, Noah Pettit, Michael Hausser, and Jakob H Macke. Inferring neural population dynamics from multiple partial recordings of the same neural circuit. In *Advances in Neural Information Processing Systems*, pages 539–547, 2013.
- [45] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 1936.
- [46] Yusuke Fujiwara, Yoichi Miyawaki, and Yukiyasu Kamitani. Modular encoding and decoding models derived from bayesian canonical correlation analysis. *Neural computation*, 25(4):979–1005, 2013.
- [47] Jacek P Dmochowski, Jason J Ki, Paul DeGuzman, Paul Sajda, and Lucas C Parra. Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, 180:134–146, 2018.
- [48] Pei Ling Lai and Colin Fyfe. Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, 10(05):365–377, 2000.
- [49] Su-Yun Huang, Mei-Hsien Lee, and Chuhsing Kate Hsiao. Kernel canonical correlation analysis and its applications to nonlinear measures of association and test of independence. *Institute of Statistical Science: Academia Sinica, Taiwan*, 2006.
- [50] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International conference on machine learning*, pages 1247–1255, 2013.

- [51] Chethan Pandarinath, Daniel J O’Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, page 1, 2018.
- [52] Jude F Mitchell, Kristy A Sundberg, and John H Reynolds. Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron*, 63(6):879–888, 2009.
- [53] Marlene R Cohen and John HR Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594, 2009.
- [54] Ying Zhang, Ethan M Meyers, Narcisse P Bichot, Thomas Serre, Tomaso A Poggio, and Robert Desimone. Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences*, 108(21):8850–8855, 2011.
- [55] Carrie J McAdams and John HR Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, 19(1):431–441, 1999.
- [56] Vadim S Polikov, Patrick A Tresco, and William M Reichert. Response of brain tissue to chronically implanted neural electrodes. *Journal of neuroscience methods*, 148(1):1–18, 2005.
- [57] Warren M Grill, Sharon E Norman, and Ravi V Bellamkonda. Implanted neural interfaces: biochallenges and engineered solutions. *Annual review of biomedical engineering*, 11:1–24, 2009.
- [58] Douglas McCreery, Victor Pikov, and Philip R Troyk. Neuronal loss due to prolonged controlled-current stimulation with chronically implanted microelectrodes in the cat cerebral cortex. *Journal of neural engineering*, 7(3):036005, 2010.
- [59] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [60] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [61] Saskia EJ de Vries, Jerome A Lecoq, Michael A Buice, Peter A Groblewski, Gabriel K Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman,

- et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151, 2020.
- [62] Joshua H Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Gradis, Gregory Heller, Tamina K Ramirez, Hannah Choi, Jennifer A Luviano, et al. A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. *bioRxiv*, page 805010, 2019.
- [63] Chethan Pandarinath, K Cora Ames, Abigail A Russo, Ali Farshchian, Lee E Miller, Eva L Dyer, and Jonathan C Kao. Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. *Journal of Neuroscience*, 38(44):9390–9401, 2018.
- [64] Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2765–2781, 2013.
- [65] Eva L Dyer, Aswin C Sankaranarayanan, and Richard G Baraniuk. Greedy feature selection for subspace clustering. *The Journal of Machine Learning Research*, 14(1):2487–2517, 2013.
- [66] Ehsan Elhamifar and René Vidal. Sparse manifold clustering and embedding. In *Advances in neural information processing systems*, pages 55–63, 2011.
- [67] William E Bishop and M Yu Byron. Deterministic symmetric positive semidefinite matrix completion. In *Advances in Neural Information Processing Systems*, pages 2762–2770, 2014.
- [68] Guangcan Liu, Qingshan Liu, and Xiaotong Yuan. A new theory for matrix completion. In *Advances in Neural Information Processing Systems*, pages 785–794, 2017.