Representational sparsity determines representational stability in sensory cortices

Shanshan Qin[†] **(ssqin@seas.harvard.edu), Cengiz Pehlevan (cpehlevan@seas.harvard.edu)**
John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.
Center for Brain Science, Harvard University, Cambridge, MA 02138, USA.
[†] Current address: Center for Computational Neuroscience, Flatiron Institute, New York, NY 10010, USA.

Abstract

Recent advancements in large-scale neural activity recordings have revealed a continuous evolution in neural population activity associated with familiar tasks, percepts, and actions over extended periods. The underlying mechanisms and functional implications of such "representational drift" remain poorly understood. In many sensory cortices, representation stability varies with stimulus type. For example, in the mouse primary visual cortex, natural movie stimuli induce drift, unlike drifting gratings. To understand the mechanism behind such stimulus-dependent representational drift in visual cortex, we propose that natural stimuli prompt denser responses compared to artificial ones, making denser representations more susceptible to synaptic noise. We evaluated this hypothesis by training a sparse coding network with continually updating synaptic weights. We found that representations for more complex image patches are denser and also exhibit more drift compared to simpler ones. This result is consistent with experimental findings. To further explore the relationship between drift speed and representational sparsity, we developed a mean-field model to analyze how different noise sources contribute to drift. Our model provides a plausible explanation for stimulus-dependent representational drift.

Keywords: Representational drift; sparsity; visual cortex

Introduction

The spatial receptive fields of simple cells in primary visual cortex are characterized by localized, oriented Gabor-like filters. A classic account for the formation of such receptive fields is the sparse coding model (Olshausen & Field,). This model postulates that neural networks in the visual cortex has evolved to efficiently represent natural images in the sense that each image only triggers a few neurons to respond. Mathematically, this can be formulated as an unsupervised learning task with the following objective function

$$\min_{\mathbf{W}} \sum_{t}^{T} \left(\min_{\mathbf{y}_t} \frac{1}{2} \|\mathbf{x_t} - \mathbf{W}^\top \mathbf{y}_t\|_2^2 + \lambda \|\mathbf{y}_t\|_1 \right), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the vectorized image pixels, $\mathbf{y}_t \in \mathbb{R}^d$ is the output neuronal response population vector, each row of $\mathbf{W} \in \mathbb{R}^{d \times n}$ is a basis function. The objective function aims to find a set of basis functions with sparse coefficients $\{y_i\}$ that can best reconstruct the image $\mathbf{x}_t$. Optimizing (1) leads to a network with forward and recurrent connections, and the locally competitive algorithm (LCA) (Rozell, Johnson, Baraniuk, & Olshausen,). The neural dynamics is then described as

$$\frac{du_i}{dt} = -u_i + \sum_j W_{ij} x_j - \sum_{j \neq i} L_{ij} y_j, \tag{2}$$

$$y_i = g(u_i) = \text{sign}(u_i)[|u_i| - \lambda]_+ \tag{3}$$

and the learning rule is

$$\Delta W_{ij} = \eta \left(y_i x_j - y_i \sum_{k=1}^{d} y_k W_{kj} \right). \tag{4}$$

Here u_i is the membrane potential of i-th output neuron, $\mathbf{L} \equiv \mathbf{W}\mathbf{W}^\top$. Each row of $\mathbf{W}$ are then normalized after each update step. When trained on natural image patches, this network evolves to a configuration where each output neuron has Gabor-like receptive field (Fig.1). With continual online learning, these receptive fields fluctuate, leading to the drift of representations.

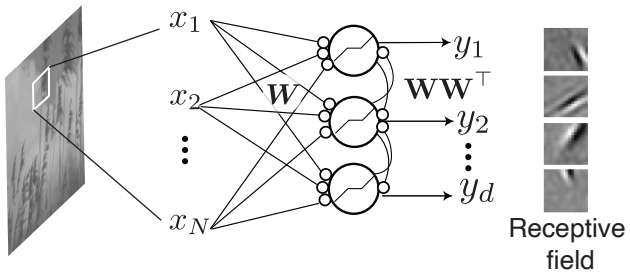

Figure 1: Locally competitive algorithm for sparse coding in visual cortex.

Representational drift speed is correlated with sparsity. In our study, the neural network was first trained at batch mode to a stationary state (Olshausen & Field,). The receptive field of each output neuron corresponds to each row of $\mathbf{W}$, denoted as ϕ_i^* and it is also called "filter" or basis function. The input consists of whitened 16×16 image patches. Subsequently, we conduct online training, where image patches are presented to the network one at a time. We focused on assessing the impact of synaptic noise in this ongoing online learning on the representations of different "probe" image patches (Fig. 2A): (1) Filters from the initial stage, i.e, ϕ_i^*; (2) Static gratings; (3) Training image patches; (4) Novel natural image patches. We quantified the drift speed by the population vector (PV) $\mathbf{y}(t)$ using its autocorrelation coefficient. Our findings reveal that natural image patches (used in training or novel) drift more rapidly compared to simpler patches (learned receptive fields and gratings) (Fig. 2B). Interestingly, natural image patches also have denser representations as quantified by the fraction of average active output neurons (Fig. 2C). This suggests that random synaptic noise due to online learning has stronger effect on the stability of denser representations.

To further examine sparsity-dependent drift speed, we assessed the representational stability of a set of image patches with increasing complexity, i.e., $\hat{\mathbf{x}} = \sum_{i=1}^{K} a_i \phi_i^*$, where ϕ_i^* is the i-th row of the stationary $\mathbf{W}$ (Fig. 3A). A close exam of the evolution of $\mathbf{W}$ under online learning showed that its fluctuation can be approximated as an O-U process

$$d\delta W_{ij}(t) = -\delta W_{ij}(t)dt + \sqrt{2Ddt}\xi_{ij}(t), \tag{5}$$

where D is the diffusion matrix and $\xi_{ij}(t)$ is the standard white Gaussian noise. Indeed, we see the representations

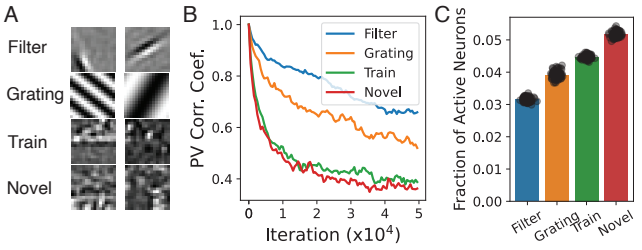

Figure 2: Drift of population vectors for four different type of input (A) under continual update of $\mathbf{W}$ as quantified by the autocorrelation coefficient of population vector (B). Representational sparsity of different probe patches (C).

of more complex image patches drift faster compared with simpler patches under the dynamics of (5)(Fig. 3B). Meanwhile, more output neurons are active for more complex image patches (Fig. 3C). To quantify the drift speed, we fit an exponential decay to the PV correlation coefficients, the plateau residual correlation coefficient showed a clear dependence on the representational sparsity (Fig. 3D).

Interestingly, when we train a multi-layer perceptron to reconstruct input image patches using stochastic gradient descent algorithm injected with Gaussian noise, we do not observe the above sparsity-dependent drift speed phenomenon (Fig. 4). In our ongoing study, we are

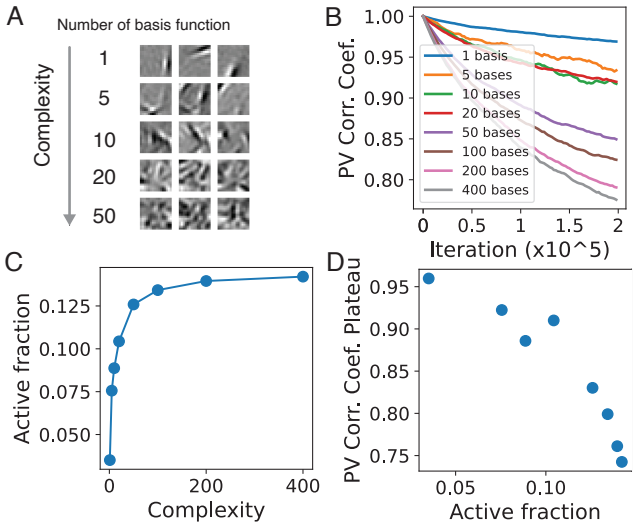

Figure 3: Under the perturbation of $\mathbf{W}$ (5), more complex synthetic image patches (A) drift faster as quantified by the autocorrelation coefficients of PV (B). Meanwhile, more complex patches have denser representations (C) and smaller plateau PV correlation coefficients (D).

Representational sparsity of primary visual cortex neurons. We next examine the representational sparsity of primary visual cortex neurons to different sensory stimuli, such as static gratings, natural scenes, drifting gratings and natural movies. Overall, we found a clear dependence of representational sparsity on the complexity of stimuli (Fig. 5).

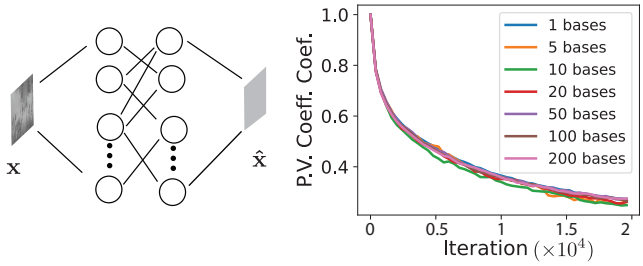

Figure 4: Training a MLP to reconstruct input image patches with SGD noise. Representations of different synthetic image patches drift with similar speed.

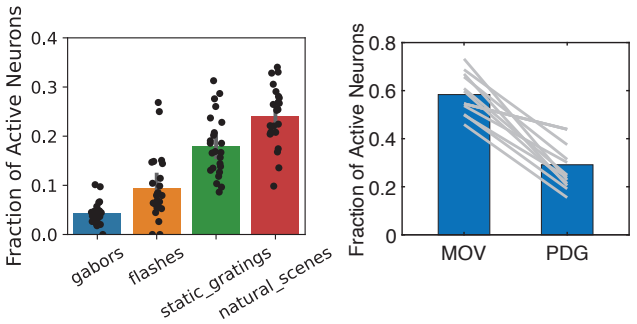

Figure 5: Representational sparsity of different visual stimuli in mouse primary visual cortex. Left: fraction of active neurons when presented with different visual stimuli. Data from (de Vries et al.,). Right: fraction of active neurons when responding to natural movie and passive drifting gratings. Data from (Marks & Goard,).

Discussion

We show that input-dependent drift could be captured by continual learning in a sparse coding network model with synaptic noise. Given the different representational sparsity for natural and artificial stimuli, the ongoing synaptic noise has different effects. Overall, denser representations are more sensitive to the noise. Our comparison with a MLP trained with SGD algorithm indicates that input-dependent drift is a not generic features of artificial neural networks (ANNs) trained for image processing. This assertion will be examined by more systematic studies on ANNs with different architectures.

References

de Vries, S. E., Lecoq, J. A., Buice, M. A., Groblewski, P. A., Ocker, G. K., Oliver, M., ... others (2020). A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature neuroscience*, *23*(1), 138–151.

Marks, T. D., Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature communications*, *12*(1), 5169.

Olshausen, B. A., Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. , *381*, 607-609. doi: 10.1038/381607a0

Rozell, C. J., Johnson, D. H., Baraniuk, R. G., Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, *20*(10), 2526–2563.