CellPress

Computation through Cortical Dynamics

Laura N. Driscoll,[1] Matthew D. Golub,[1] and David Sussillo[1,2,3,*]
[1]Department of Electrical Engineering, Stanford University, Stanford, CA, USA
[2]Stanford Neurosciences Institute, Stanford University, Stanford, CA, USA
[3]Google Brain, Google, Inc., Mountain View, CA, USA
*Correspondence: sussillo@google.com
https://doi.org/10.1016/j.neuron.2018.05.029

Population dynamics is emerging as a language for understanding high-dimensional neural recordings. Remington et al. (2018) explore how inputs to frontal cortex modulate neural dynamics in order to implement a computation of interest.

The original conception of reservoir computing (Lukoševičius et al., 2012) came with an image of throwing a pebble into a pond. Almost any dynamical system, in this case the three-dimensional physics of water in the pond, can be used for computation or memory. For example, by examining the current state of the concentric waves created by a pebble falling into a pond at an earlier time, one can determine when and where the pebble initially hit the water. The reservoir computing literature is often concerned with computing through the dynamics of a fixed physical medium (e.g., optical computers). But if one could optimally specify a dynamical system that was tailor-made for a specific computation or a set of related computations, what would the dynamics of that system look like? For example, what if one could change the physics of the water to optimally subserve the task of remembering the time when the pebble dropped into the water? How would such a dynamical system be configured?

The motivation for these questions is rooted in a growing body of experimental and theoretical work (Carnevale et al., 2015; Churchland et al., 2012; Kato et al., 2015; Machens et al., 2005; Mante et al., 2013; Mazor and Laurent, 2005) that argues neural dynamics implement computation. The dynamics are those of the neural population state (e.g., binned spike counts) evolving through time. According to this "dynamical systems hypothesis," the dynamics describe how a neural population moves between key states. For example, a neural population may move from an initial state that represents a question to a later state that represents the answer to that question, and the transition of the neural population from initial to final state implements the computation. Input to the circuit reconfigures the dynamics and thereby specifies the kind of answer desired. We are thus describing a "computation *through* dynamics."

Mathematically, the dynamical systems hypothesis says that the neural population state can be described by $\dot{x}(t) = F(x(t), u(t))$, with $x(t)$ the neural population state, $u(t)$ an input to the system, and $F(\cdot)$ a nonlinear function. This equation asserts that any change in the state of the system can be calculated from the current state of the neural population and the inputs to that population. The initial conditions, $x(0)$, and the input, $u(t)$, are combined in the nonlinear function $F(\cdot)$ to determine the state of the system as it updates through time. Note that $u(t)$ has to be reasonably constrained; otherwise, the equation is not particularly useful as a dynamical model because $u(t)$ could completely drive the system in arbitrary ways. In summary, a neural computation could be configured through dynamics by specifying both the initial condition, $x(0)$, and the input, $u(t)$.

Remington et al. (2018) pose the key question of how these initial conditions and inputs are configured in dorsal medial frontal cortex (DMFC) during a timing task. To probe the role of DMFC, the Jazayeri lab uses an experimental paradigm called "Ready, Set, Go" (RSG). RSG requires the subject to track and reproduce various timing intervals. In this task, there are two successive stimuli for "Ready" and "Set," after which the subject must respond "Go," following a target wait interval. Two key independent variables of this task are the initial time interval between Ready and Set, called the sample interval, t_s, and the desired delay between Set and Go, called the target interval, t_t. The authors introduced seven conditions to the task by allowing seven values for t_s. They further defined two contexts by defining t_t as $t_t = g\ t_s$, with g taking values of either 1 or 1.5. According to the dynamical systems hypothesis, t_s and g must enter the system through an input to the system, as an initial condition, or a combination of both. The way in which these independent variables are fed into and represented by cortex results in a particular organization of neural trajectories for trials of this task. It is this organization of the resultant neural trajectories that Remington et al. (2018) study.

There are complexities to studying high-dimensional neural population state trajectories. Imagine watching a play from the far wing of the theater, where you can only see the shadows of the actors projected onto the stage. You may glean broad understanding of the play, but your comprehension of the performance will undoubtedly be impoverished. So it is with neural trajectories, which are almost always high-dimensional, even when accounting for correlations among neurons. Modern technology such as multi-unit electrode arrays or calcium imaging now allows us to record higher-dimensional neural activity than ever before. Yet our capacity to visualize is limited by our three-dimensional physical reality. This disparity means that the common technique of projecting data down to three dimensions for visualization is fundamentally problematic because just as if we attempt to understand the play from the shadows cast onto the stage, visualizing low-dimensional

projections of data can miss key features of neural activity.

There are additional concerns when interpreting low-dimensional visualizations of neural data. We tend to think in idealized terms of straight lines in a noise-free environment, but neural data are not ideal. Neural trajectories are often curved and contain variability not readily explained by external covariates. Finally, techniques that use optimization to discover specific low-dimensional projections of interest must be used carefully; otherwise, they may find projections consistent with nearly any hypothesis. These problematic features of low-dimensional projections make it difficult to use statistics to quantify the geometry of projections in two- and three-dimensional spaces. Instead, one must find the appropriate number of dimensions to capture the majority of neural variance (typically 90% or greater) and perform statistics in these high-dimensional spaces (Afshar et al., 2011; Ames et al., 2014). We then have three relevant spaces to work in: (1) two- or three-dimensional space for visualization; (2) high-dimensional "denoised" space for statistics that explains the majority of neural variance, typically 10- to 20-dimensional; and (3) full dimensional space of all recorded units. Below, we use high-dimensional to refer to the last two cases.

If we really believe that the geometry of high-dimensional neural data has meaning in terms of both behavior and the computations that subserve that behavior, then we are obligated to move beyond three-dimensional projections and attempt to rigorously quantify high-dimensional neural geometry. In this work, Remington et al. (2018) developed a set of tools, called kinematic analysis of neural trajectories (KiNeT), that is used to measure features of trajectories in high-dimensional population-activity space. KiNeT measures timing, distances, and angles between neural trajectories. The team used KiNeT to analyze these trajectories in their timing task to make inferences about the dynamical structure of the neural state space underlying the task.

Using KiNeT, the authors found that the neural dynamics in frontal cortex for this timing task were organized in a modular structure such that the sample interval (t_s) and gain (g) variables were encoded on different axes in neural state space. The flow of all neural trajectories through time moved along roughly parallel paths that were essentially orthogonal to axes encoding the sample interval and gain. The authors were able to rigorously quantify this geometry using KiNeT. The authors also excluded four additional hypothetical organizations of the neural trajectories. Finally, they showed that location in neural state space had behavioral implications. For example, the average across neural trajectories with a given sample interval that were closest in neural state space to trajectories subserving longer sample intervals was associated with trials that were also slower behaviorally on average. We emphasize that the authors were working in a high-dimensional space created by the ten principal components that explained the most variance of the neural data.

To further their study, the authors then used trained recurrent neural networks (RNNs) for hypothesis building. With the growth of the computation through dynamics hypothesis, RNNs have gained traction as learnable, high-dimensional dynamical models that can explicitly test hypotheses about how neural dynamics might be structured when subserving a computation (Carnevale et al., 2015; Mante et al., 2013; Sussillo et al., 2015). Broadly, the RNN is trained to perform a task analogous to the one the animal is required to perform. Concretely, given particular inputs (stimuli), the RNN will produce a corresponding target (a decision or motor output), which is designed by the modeler. After training, the RNN will produce novel outputs for novel inputs that were not used in training. The real power of the trained RNN is that it provides the modeler with a generative nonlinear dynamical system that can be compared to the neural data. Because modelers have full access to a dynamical system that performs the task of interest (e.g., unlimited trial counts and novel trial types, and access to the activations of all units in the network and all model parameters), trained RNNs can be extremely useful for hypothesis generation and for developing intuitions about the underlying computational dynamics that may subserve neural data.

Remington et al. (2018) created a synthetic version of the RSG task for the RNNs. In order to test their hypothesis of how the two independent variables, stimulus time and gain, might influence the computation in DFMC, they trained two RNN variants on this synthetic RSG task. The authors varied whether gain came into the RNN as a transient input, thus acting to influence only the initial condition of the RNN, or as a static, "contextual" input that remained on during the entirety of the trial. The two RNN variants were then compared to the neural data both visually and using KiNeT. Geometrically, the neural data were quantitatively more similar to the RNN variant where g entered the RNN as a static input. The authors thus argued that the RNN studies provide evidence in support of the hypothesis that g enters the neural circuit as a static contextual input, which is also consistent with how the visual stimulus representing g was delivered to the animal.

The work of Remington et al. (2018) presents an opportunity to make two broader points. First, the hypothesis that contextual information is represented in neural activity as a static input in the RSG task suggests that cortical input may flexibly reconfigure local cortical circuits. This contextual input shifts the neural activity in state space, thereby changing the effective circuit by shifting each neuron's average operating point on its nonlinearity. In this manner, a contextual input may allow the same group of cortical neurons to perform related computations through a mild reconfiguration of the dynamics rather than requiring novel circuits for each new task (see also Mante et al., 2013). This result could have major implications for how biological neural networks are able to generalize across related tasks with minimal task-specific training.

Second, trained RNNs can be extremely useful for hypothesis generation and for developing intuitions about the computations that implement various algorithms. However, caution must be used in drawing strong conclusions about the relationship between an RNN and neural data. An RNN merely provides one example of how the computation could be solved and does not prove that the brain solves computations in the same way. To gain the most meaningful insight from RNN models, we as a community should strive to make our RNN

models as precise as possible by iterating through multiple input/output configurations, hyper-parameters such as initialization schemes, and other augmentations such as regularization methods. These choices, which are not optimized, can have profound effects on the dynamical solution learned by an RNN. We should also strive to orient our work with RNNs to make testable predictions that can be experimentally examined. In this way, the systems community will be able to more effectively utilize trained RNNs to support or refute hypotheses about neural computation in biological neural circuits.

In summary, Remington et al. (2018) advances the field by placing computation through dynamics squarely at the center of neuroscientific investigation and further provides a set of quantitative tools to rigorously interrogate the geometry of high-dimensional neural dynamics. As we move forward, cleverly chosen two- and three-dimensional visualizations will continue to be indispensable in developing basic intuitions about the structure of population activity and the computations such structure subserves. However, with new experimental methodologies that allow increasingly large population recordings, and with the development of high-complexity behaviors, quantification tools amenable to high-dimensional spaces will become increasingly relevant.

REFERENCES

Afshar, A., Santhanam, G., Yu, B.M., Ryu, S.I., Sahani, M., and Shenoy, K.V. (2011). Single-trial neural correlates of arm movement preparation. Neuron *71*, 555–564.

Ames, K.C., Ryu, S.I., and Shenoy, K.V. (2014). Neural dynamics of reaching following incorrect or absent motor preparation. Neuron *81*, 438–451.

Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic control of response criterion in premotor cortex during perceptual detection under temporal uncertainty. Neuron *86*, 1067–1077.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Kato, S., Kaplan, H.S., Schrödel, T., Skora, S., Lindsay, T.H., Yemini, E., Lockery, S., and Zimmer, M. (2015). Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. Cell *163*, 656–669.

Lukoševičius, M., Jaeger, H., and Schrauwen, B. (2012). Reservoir computing trends. KI-Künstliche Intelligenz *26*, 365–371.

Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science *307*, 1121–1124.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Mazor, O., and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron *48*, 661–673.

Remington, E.D., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. Neuron *98*, this issue, 1005–1019.

Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

Hearing out Ultrasound Neuromodulation

Raag D. Airan[1,*] and Kim Butts Pauly[1,2,3,*]
[1]Department of Radiology, Stanford University, Stanford, CA 94305, USA
[2]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
[3]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
*Correspondence: rairan@stanford.edu (R.D.A.), kimbutts@stanford.edu (K.B.P.)
https://doi.org/10.1016/j.neuron.2018.05.031

Many neuroscientists are excited regarding the potential of ultrasound to yield spatiotemporally precise and noninvasive modulation of arbitrary brain regions. Here, Guo et al. (2018) and Sato et al. (2018) show that applying ultrasound to rodent brains activates acoustic responses more prominently than eliciting neuromodulation directly, suggesting potential confounds of ultrasound neuromodulation experiments.

Current techniques for noninvasive neuromodulation, such as transcranial magnetic stimulation (TMS) or either transcranial alternating or direct current stimulation (tcACS or tcDCS), show a limiting trade-off between the spatial resolution and depth of penetration of the intervention (Deng et al., 2013). In contrast, focused ultrasound can deliver energy with millimeter-scale spatial resolution to any point of the brain, with guidance and real-time visualization of the ultrasound focus with MRI, using hardware that is already clinically available (Elias et al., 2016; Hynynen and Clement, 2007). Since as early as the 1950's, electrophysiological, functional neuroimaging, and behavioral effects have been reported after applying focused ultrasound to the mammalian brain across a range of species, including mice, rats, cats, monkeys, and humans (reviewed in depth in Tyler et al., 2018). These features and data have led to a surge of recent interest in developing focused ultrasound as a tool for noninvasive neuromodulation. However, the mechanism by which ultrasound may interact with neural tissue to drive these effects, as well as the robustness of this mechanism, has been